

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:44:48 ; Search time 0.585209 Seconds
(without alignments)
292.168 Million cell updates/sec

Title: US-09-606-129A-18

Perfect score: 41

Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	78.0	40	1	US-07-822-966B-4
2	32	78.0	88	1	US-08-336-618-19
3	32	78.0	99	5	PCT-US92-03993-5
4	32	78.0	141	1	US-07-822-966B-6
5	32	78.0	141	2	US-08-803-899-6
6	32	78.0	142	1	US-08-336-618-18
7	32	78.0	142	5	PCT-US92-03993-7
8	32	78.0	269	4	US-09-006-783A-5
9	32	78.0	294	2	US-09-258-371-10
10	32	78.0	294	3	US-08-751-230-10
11	32	78.0	294	4	US-09-499-082-10
12	32	78.0	294	4	US-09-258-372-10
13	32	78.0	294	4	US-09-006-783A-3
14	32	78.0	527	4	US-09-659-166-2
15	32	78.0	528	2	US-08-802-466-2
16	32	78.0	528	4	US-09-350-484-2
17	32	78.0	565	2	US-08-892-770-6
18	32	78.0	571	2	US-08-892-770-5
19	31	75.6	504	3	US-09-231-529-6
20	31	75.6	504	4	US-08-977-816-6
21	31	75.6	550	3	US-09-039-859-9
22	30	73.2	508	2	US-08-818-024-3
23	30	73.2	508	4	US-09-334-775A-3
24	29	70.7	257	2	US-08-715-204-6
25	29	70.7	257	3	US-09-162-597-6
26	29	70.7	396	3	US-09-082-310-2
27	29	70.7	568	2	US-08-835-170-4

28	29	70.7	568	4	US-09-359-257-4	Sequence 4, Appli
29	29	70.7	588	2	US-08-835-170-2	Sequence 2, Appli
30	29	70.7	588	4	US-09-359-257-2	Sequence 2, Appli
31	28	68.3	22	1	US-08-484-635-166	Sequence 166, App
32	28	68.3	22	2	US-08-484-631-166	Sequence 166, App
33	28	68.3	22	2	US-08-827-570-166	Sequence 166, App
34	28	68.3	55	1	US-08-469-427A-3	Sequence 3, Appli
35	28	68.3	55	2	US-08-609-443B-3	Sequence 3, Appli
36	28	68.3	55	2	US-08-569-063C-3	Sequence 3, Appli
37	28	68.3	188	1	US-08-469-427A-5	Sequence 5, Appli
38	28	68.3	188	2	US-08-609-443B-5	Sequence 5, Appli
39	28	68.3	188	2	US-08-569-063C-5	Sequence 5, Appli
40	28	68.3	195	1	US-08-469-427A-7	Sequence 7, Appli
41	28	68.3	195	2	US-08-609-443B-7	Sequence 7, Appli
42	28	68.3	195	2	US-08-569-063C-7	Sequence 7, Appli
43	28	68.3	477	2	US-08-560-098A-51	Sequence 51, Appli
44	28	68.3	873	3	US-08-990-140-2	Sequence 2, Appli
45	28	68.3	873	4	US-09-546-238-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-07-822-966B-4.
; Sequence 4, Application US/07822966B
; Patent No. 5498597
; GENERAL INFORMATION:
; APPLICANT: Steven J. Burakoff
; APPLICANT: Stuart L. Schreiber
; APPLICANT: Barbara E. Blier
; TITLE OF INVENTION: FKBP-13, AN FK506-BINDING
; TITLE OF INVENTION: IMMUNOPHILIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/822,966B
; FILING DATE: January 17, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/052001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-07-822-966B-4

Query Match 78.0%; Score 32; DB 1; Length 40;
Best Local Similarity 71.4%; Pred No. 3.8;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KKRIMHC 7
Db 14 KKRVDHC 20

RESULT 2
US-08-336-618-19
; Sequence 19, Application US/08336618
; Patent No. 5763590
; GENERAL INFORMATION:
; APPLICANT: Peattie, Debra A.
; APPLICANT: Harding, Matthew W.
; APPLICANT: Livingston, David J.
; TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
; TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
; TITLE OF INVENTION: CDNA
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,618
; FILING DATE: 09-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/963,325
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/777,752
; FILING DATE: 11-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/
; FILING DATE: 09-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: VPI91-06A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-618-19

Query Match 78.0%; Score 32; DB 1; Length 88;
Best Local Similarity 71.4%; Pred. No. 8.1;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
Db 14 KKRVDHC 20

RESULT 3
PCT-US92-03993-5
; Sequence 5, Application PC/TUS9203993
; GENERAL INFORMATION:
; APPLICANT: Harding, Matthew W.
; TITLE OF INVENTION: FKBP: A NOVEL PROLYL ISOMERASE AND
; TITLE OF INVENTION: RAPAMYCIN/FK506 BINDING PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/822,966B
; FILING DATE: January 17, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/052001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154

STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/03993
FILING DATE: 19920507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/697,113
FILING DATE: 08-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: VPI91-05A PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 99 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-03993-5

Query Match 78.0%; Score 32; DB 5; Length 99;
Best Local Similarity 71.4%; Pred. No. 9.1;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
Db 14 KKRVDHC 20

RESULT 4
US-07-822-966B-6
; Sequence 6, Application US/07822966B
; Patent No. 5498597
; GENERAL INFORMATION:
; APPLICANT: Steven J. Burakoff
; APPLICANT: Stuart L. Schreiber
; APPLICANT: Barbara E. Bierer
; TITLE OF INVENTION: FKBP-13, AN FK506-BINDING
; TITLE OF INVENTION: IMMUNOPHILIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/822,966B
; FILING DATE: January 17, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/052001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
us-07-822-966B-6

Query Match 78.0%; Score 32; DB 1; Length 141;
Best Local Similarity 71.4%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||: ||
Db 35 KKRVDHC 41

RESULT 5
us-08-803-899-6
Sequence 6, Application US/08803899
Patent No. 5912224
GENERAL INFORMATION:
APPLICANT: DONAHOE, PATRICIA K.
APPLICANT: WANG, TONGWEN
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
CELLULAR RESPONSE TO TGF-BETA LIGANDS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/803.899
FILING DATE: 02/21/1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/012,054
FILING DATE: 02/22/1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 0609.4240001
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-803-899-6

Query Match 78.0%; Score 32; DB 2; Length 141;
Best Local Similarity 71.4%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||: ||
Db 35 KKRVDHC 41

RESULT 6
us-08-336-618-18
Sequence 18, Application US/08336618
Patent No. 5763590
GENERAL INFORMATION:
APPLICANT: Peattie, Debra A.
APPLICANT: Harding, Matthew W.
APPLICANT: Livingston, David J.
TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
CDNA
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336.618
FILING DATE: 09-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/963,325
FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/777,752
FILING DATE: 11-OCT-1991

ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: VP191-06A
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 142 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-336-618-18

Query Match 78.0%; Score 32; DB 1; Length 142;
Best Local Similarity 71.4%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||: ||
Db 36 KKRVDHC 42

RESULT 7
PCT-US92-03993-7
Sequence 7, Application PC/TUS9203993
GENERAL INFORMATION:
APPLICANT: Harding, Matthew W.
TITLE OF INVENTION: RFKBP: A NOVEL PROLYL ISOMERASE AND
RAPAMYCIN/FK506 BINDING PROTEIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA

```
;
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/03993
; FILING DATE: 19920507
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/697,113
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: VP191-05A PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; LENGTH: 142 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-03993-7

Query Match 78.0%; Score 32; DB 5; Length 142;
Best Local Similarity 71.4%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
Db 36 KKRVDHC 42

RESULT 8
US-09-006-783A-5
; Sequence 5, Application US/09006783A
; Patent No. 6297366
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrey V
; APPLICANT: Garkavstev, Igor
; APPLICANT: Riabowol, Karl
; TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
; TITLE OF INVENTION: Pathway
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/006,783A
; FILING DATE: 15-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6297366nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 97,837
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 5:
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-006-783A-5

Query Match 78.0%; Score 32; DB 4; Length 269;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
Db 67 KRRMLHC 73

RESULT 9
US-09-258-371-10
; Sequence 10, Application US/09258371
; Patent No. 5986078
; GENERAL INFORMATION:
; APPLICANT: Garkavtsev, Igor
; APPLICANT: Riabowol, Karl
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
; TITLE OF INVENTION: SUPPRESSOR GENE ING1
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/258,371
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/751,230
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mool, Leslie A.
; REGISTRATION NUMBER: 37,047
; REFERENCE/DOCKET NUMBER: 028722-144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-7400
; TELEFAX: 415-854-8275
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-258-371-10

Query Match 78.0%; Score 32; DB 2; Length 294;
Best Local Similarity 57.1%; Pred. No. 26;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
Db 82 KRRMLHC 88

RESULT 10
US-08-751-230-10
; Sequence 10, Application US/08751230
; Patent No. 6117633
```

GENERAL INFORMATION:
APPLICANT: Garkavtsev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
TITLE OF INVENTION: SUPPRESSOR GENE INGI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/751,230
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mooi, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-144
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-751-230-10

Query Match 78.0%; Score 32; DB 3; Length 294;
Best Local Similarity 57.1%; Pred. No. 26;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
I:::||
Db 82 KRRMLHC 88

RESULT 11
US-09-499-082-10
Sequence 10, Application US/09499082
Patent No. 6143522
GENERAL INFORMATION:
APPLICANT: Helbing, Caren C.
APPLICANT: Riabowol, Karl
APPLICANT: Johnston, Randall N.
APPLICANT: Garkavtsev, Igor
TITLE OF INVENTION: METHODS OF MODULATING APOPTOSIS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/499,082
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/828,158
FILING DATE: 27-MAR-1997
APPLICATION NUMBER: US 08/751230
FILING DATE: 15-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mooi, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-148
TELEPHONE: 650-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-499-082-10

Query Match 78.0%; Score 32; DB 4; Length 294;
Best Local Similarity 57.1%; Pred. No. 26;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
I:::||
Db 82 KRRMLHC 88

RESULT 12
US-09-258-372-10
Sequence 10, Application US/09258372
Patent No. 6238918
GENERAL INFORMATION:
APPLICANT: Garkavtsev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
TITLE OF INVENTION: SUPPRESSOR GENE INGI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/258,372
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/751,230
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mooi, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-144
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:

; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-258-372-10

Query Match 78.0%; Score 32; DB 4; Length 294;
Best Local Similarity 57.1%; Pred. No. 26;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
|:|:|:|
Db 82 KRRMLHC 88

RESULT 13
US-09-006-783A-3
; Sequence 3, Application US/09006783A
; Patent No. 6297366
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrey V
; APPLICANT: Garkavstev, Igor
; APPLICANT: Riabowol, Karl
; TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
; TITLE OF INVENTION: Pathway
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/006,783A
; FILING DATE: 15-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6297366nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 97,837
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-006-783A-3

Query Match 78.0%; Score 32; DB 4; Length 294;
Best Local Similarity 57.1%; Pred. No. 26;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
|:|:|:|
Db 82 KRRMLHC 88

RESULT 14
US-09-659-166-2
; Sequence 2, Application US/09659166
; Patent No. 6355465
; GENERAL INFORMATION:
; APPLICANT: CREASY, CARETHA LEE
; APPLICANT: TESTA, TANIA TAMSON

; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30218
; CURRENT APPLICATION NUMBER: US/09/659,166
; CURRENT FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: UK 9921505.5
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 527
; TYPE: PRT
; ORGANISM: rattus
US-09-659-166-2

Query Match 78.0%; Score 32; DB 4; Length 527;
Best Local Similarity 71.4%; Pred. No. 46;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
|:|:|:|
Db 267 KNRIHC 273

RESULT 15
US-08-802-466-2
; Sequence 2, Application US/08802466
; Patent No. 5972606
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,466
; FILING DATE: 19 February 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 528 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-802-466-2

Query Match 78.0%; Score 32; DB 2; Length 528;
Best Local Similarity 71.4%; Pred. No. 46;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
|:|:|:|

Wed Nov 6 14:27:41 2002

us-09-606-129a-18.ra1

Page 7

Db 268 KNRIHC 274

Search completed: November 1, 2002, 14:51:57
Job time : 2.58521 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:40:53 ; Search time 1.58682 Seconds
(without alignments)
489.985 Million cell updates/sec

Title: US-09-606-129a-18

Perfect score: 41

Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	87.8	285	21 AAG28710	Arabidopsis thalia
2	34	82.9	94	21 AAY97243	Murine INGB1 isofo
3	34	82.9	279	21 AAY97242	Murine P37ING1 pol
4	33	80.5	508	21 AAG48146	Arabidopsis thalia
5	33	80.5	533	21 AAG48145	Arabidopsis thalia
6	33	80.5	564	21 AAG48144	Arabidopsis thalia
7	32	78.0	40	17 AAR93552	Bovine FRBP-13 imm
8	32	78.0	83	22 ABG14241	Novel human diagno
9	32	78.0	83	22 AAG76114	Human colon cancer
10	32	78.0	94	21 AAY97245	Human INGB1 isofo
11	32	78.0	99	13 AAR28979	Bovine RFBKP. Bos

12	32	78.0	104	21 AAG03758	Human secreted pro
13	32	78.0	128	20 AAY03245	Amino acid sequenc
14	32	78.0	137	22 AAB87656	Bovine mammary tis
15	32	78.0	141	17 AAR93551	Human FRBP-13 immu
16	32	78.0	141	21 AAB53247	Human colon cancer
17	32	78.0	182	22 AAU31050	Novel human secret
18	32	78.0	235	22 AAB84700	Amino acid sequenc
19	32	78.0	279	21 AAY97244	Human P37ING1. Ho
20	32	78.0	279	22 AAE06676	Tumour suppressor
21	32	78.0	279	22 AAB84698	Amino acid sequenc
22	32	78.0	294	18 AAW19119	Tumour suppressor
23	32	78.0	294	19 AAW79675	Human p37ING1 poly
24	32	78.0	294	20 AAY03244	Amino acid sequenc
25	32	78.0	294	22 AAB84697	Amino acid sequenc
26	32	78.0	328	22 AAG98906	E. coli growth and
27	32	78.0	455	22 ABB71982	Drosophila melanog
28	32	78.0	528	19 AAE64559	Human protein kina
29	32	78.0	528	22 AAE02011	Human YAK1 (hYAK1)
30	32	78.0	549	21 AAY68781	Amino acid sequenc
31	32	78.0	565	20 AAW96316	Acidic leucine ami
32	32	78.0	571	20 AAW96315	Acidic leucine ami
33	32	78.0	705	22 AAG24213	Novel human diagno
34	32	78.0	795	22 AAU29777	Novel human secret
35	31	75.6	63	22 ABG09069	Novel human diagno
36	31	75.6	82	22 AAM87522	Human immune/haema
37	31	75.6	103	22 ABG08440	Novel human diagno
38	31	75.6	109	22 ABB57919	Drosophila melanog
39	31	75.6	350	22 ABB50167	Human transcriptio
40	31	75.6	398	22 ABB62446	Drosophila melanog
41	31	75.6	504	18 AAW36140	Bovine P58 protein
42	31	75.6	504	22 ABG09790	Novel human diagno
43	31	75.6	550	19 AAW71468	Carcospora nicotia
44	31	75.6	550	21 AAY99884	Carcospora nicotia
45	31	75.6	709	22 AAM93837	Human polypeptide,

ALIGNMENTS

RESULT 1

AAG28710

ID AAG28710 standard; Protein; 285 AA.

AC AAG28710;

XX 17-OCT-2000 (first entry)

DT Arabidopsis thaliana protein fragment SEQ ID NO: 34031.

DE protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 23-APR-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 28-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 87.8%; Score 36; DB 21; Length 285;
 Best Local Similarity 71.4%; Pred. No. 21;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKRIMHC 7
 |:::|
 Db 13 KKRLLHC 19

RESULT 2
 AAY97243
 ID AAY97243 standard; Protein; 94 AA.

XX AC AAY97243;
 XX 19-DEC-2000 (first entry)
 DE Murine INGB1 isoform N-terminal peptide.
 XX p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
 KW INGL; ingl; p33ING1; p37ING1; oncogene; gene therapy; diagnosis;
 KW proliferation disorder; transformation; transformed cell; mouse.
 XX Mus musculus.
 XX WO200046370-A1.
 XX 10-AUG-2000.
 XX 04-FEB-1999; 2000WO-US02959.
 XX 04-FEB-1999; 99US-0118941.
 XX (UNII) UNIV ILLINOIS FOUND.
 PI Gudkov A, Zeremski M, Gurova KV, Grigorian IA;
 XX WPI; 2000-491278/43.
 XX Detecting nucleic acid encoding exon 1b of ingl, useful for diagnosing
 PT and treating cancer, comprises contacting sample with isolated nucleic
 PT acid comprising sequence of exon 1b and detecting hybridized products
 XX Claim 6; Fig 7a; 134pp; English.
 XX Mutations in or loss of the p53 gene occur in more than 50% of
 CC human tumours and tumour cell lines, but functional inactivation of
 CC the p53 pathway occurs in a much larger proportion of tumours. In
 CC many cases the mechanism of functional inactivation of the p53 gene
 CC remains unknown but p53 has been found to act in cooperation with

CC INGL. Functional cooperation between INGL and p53 suggested that
 CC INGL encoded a tumour suppressor protein that functioned within the
 CC p53 pathway. This data suggested a possible role for INGL in head
 CC and neck cancers and chromosomal location of the INGL placed it
 CC within a region that is frequently rearranged in head and neck
 CC cancers. Large scale analysis of tumours involving INGL has not
 CC revealed mutations in INGL nor significant variations in its
 CC expression suggesting that INGL was not a useful gene to study in
 CC cancer etiology. However, alternative initiation exons of the ingl
 CC gene, each having their own promoter have been discovered.
 CC Expression of one promoter (1a) produces a protein identical to
 CC INGL. Expression of a second promoter (1b) produces a protein having
 CC an identical C-terminal fragment to INGL but an additional 104
 CC N-terminal amino acids. The newly discovered protein has been
 CC designated p37ING1 (Wild type: p33ING1). p37ING1 has the
 CC characteristics of an oncogene. When overexpressed in cells (even
 CC those expressing wild type p53) p37ING1 is able to cause
 CC proliferation or transformation of those cells. Thus detecting a
 CC nucleic acid encoding exon 1b of ingl by hybridisation with an
 CC isolated nucleic acid having the sequence of exon 1b of ingl
 CC or its antisense sequence can identify individuals expressing the
 CC oncogenic form of ingl. Novel peptide sequences taken from the 104
 CC N-terminal peptide of p37ING1 can also be used to raise antibodies
 CC that can also be used in detection methods for the p37ING1 variant.
 CC The polypeptides may be useful in gene therapy for treatment of cell
 CC proliferation disorders, especially cancers and for diagnosing and
 CC studying cancers.

XX Sequence 94 AA;

Query Match 82.9%; Score 34; DB 21; Length 94;
 Best Local Similarity 57.1%; Pred. No. 18;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKRIMHC 7
 |:::|
 Db 67 KRRVLHC 73

RESULT 3
 AAY97242
 ID AAY97242 standard; Protein; 279 AA.

XX AC AAY97242;
 XX 19-DEC-2000 (first entry)
 DE Murine P37ING1 polypeptide.
 DE p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
 KW INGL; ingl; p33ING1; p37ING1; oncogene; gene therapy; diagnosis;
 KW proliferation disorder; transformation; transformed cell; mouse.
 XX Mus musculus.
 XX WO200046370-A1.
 XX 10-AUG-2000.
 XX 04-FEB-2000; 2000WO-US02959.
 XX 04-FEB-1999; 99US-0118941.
 XX (UNII) UNIV ILLINOIS FOUND.
 PI Gudkov A, Zeremski M, Gurova KV, Grigorian IA;
 XX WPI; 2000-491278/43.
 XX N-PSDB; AAA53790.

XX Detecting nucleic acid encoding exon 1b of ingl, useful for diagnosing
 PT and treating cancer, comprises contacting sample with isolated nucleic
 PT acid comprising sequence of exon 1b and detecting hybridized products

XX Disclosure; Fig 12; 134pp; English.

PS Mutations in or loss of the p53 gene occur in more than 50% of

XX human tumours and tumour cell lines, but functional inactivation of

CC the p53 pathway occurs in a much larger proportion of tumours. In

CC many cases the mechanism of functional inactivation of the p53 gene

CC remains unknown but p53 has been found to act in cooperation with

CC INK1. Functional cooperation between INK1 and p53 suggested that

CC INK1 encoded a tumour suppressor protein that functioned within the

CC p53 pathway. This data suggested a possible role for INK1 in head

CC and neck cancers and chromosomal location of the INK1 placed it

CC within a region that is frequently rearranged in head and neck

CC cancers. Large scale analysis of tumours involving INK1 has not

CC revealed mutations in INK1 nor significant variations in its

CC expression suggesting that INK1 was not a useful gene to study in

CC cancer etiology. However, alternative initiation exons of the INK1

CC gene, each having their own promoter have been discovered.

CC Expression of one promoter (1a) produces a protein identical to

CC INK1. Expression of a second promoter (1b) produces a protein having

CC an identical C-terminal fragment to INK1 but an additional 104

CC N-terminal amino acids. The newly discovered protein has been

CC designated p37ING1 (Wild type: p37ING1). p37ING1 has the

CC characteristics of an oncogene. When overexpressed in cells (even

CC those expressing wild type p53) p37ING1 is able to cause

CC proliferation or transformation of those cells. Thus detecting a

CC nucleic acid encoding exon 1b of INK1 by hybridisation with an

CC isolated nucleic acid having the sequence of exon 1b of INK1

CC or its antisense sequence can identify individuals expressing the

CC oncogenic form of INK1. Novel peptide sequences taken from the 104

CC N-terminal peptide of p37ING1 can also be used to raise antibodies

CC that can also be used in detection methods for the p37ING1 variant.

CC The polypeptides may be useful in gene therapy for treatment of cell

CC proliferation disorders, especially cancers and for diagnosing and

CC studying cancers.

XX

SQ Sequence 279 AA;

Query Match 82.9%; Score 34; DB 21; Length 279;

Best Local Similarity 57.1%; Pred. No. 51;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKRIMHC 7

I:::II

Db 67 KRRVLHC 73

RESULT 4

AA048146

ID AAG48146 standard; Protein; 508 AA.

XX

AC AAG48146;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 60770.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

XX 25-FEB-2000; 2000EP-0301439.

XX

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 06-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 04-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 13-AUG-1999;	99US-0148565.		
PR 13-AUG-1999;	99US-0148684.		
PR 16-AUG-1999;	99US-0149368.		
PR 17-AUG-1999;	99US-0149175.		
PR 18-AUG-1999;	99US-0149426.		
PR 20-AUG-1999;	99US-0149722.		
PR 20-AUG-1999;	99US-0149723.		
PR 20-AUG-1999;	99US-0149929.		
PR 23-AUG-1999;	99US-0149902.		
PR 23-AUG-1999;	99US-0149930.		
PR 25-AUG-1999;	99US-0150566.		
PR 26-AUG-1999;	99US-0150884.		
PR 27-AUG-1999;	99US-0151065.		
PR 27-AUG-1999;	99US-0151066.		
PR 27-AUG-1999;	99US-0151080.		
PR 30-AUG-1999;	99US-0151303.		
PR 31-AUG-1999;	99US-0151438.		
PR 01-SEP-1999;	99US-0151930.		
PR 07-SEP-1999;	99US-0152363.		
PR 10-SEP-1999;	99US-0153070.		
PR 13-SEP-1999;	99US-0153758.		
PR 15-SEP-1999;	99US-0154018.		
PR 16-SEP-1999;	99US-0154039.		
PR 20-SEP-1999;	99US-0154779.		
PR 22-SEP-1999;	99US-0155139.		
PR 23-SEP-1999;	99US-0155486.		
PR 24-SEP-1999;	99US-0155659.		
PR 28-SEP-1999;	99US-0156458.		
PR 29-SEP-1999;	99US-0156596.		
PR 04-OCT-1999;	99US-0157117.		
PR 05-OCT-1999;	99US-0157753.		
PR 06-OCT-1999;	99US-0157865.		
PR 07-OCT-1999;	99US-0158029.		
Query Match 80.5%; Score 33; DB 21; Length 508;			
Best Local Similarity 57.1%; Pred. No. 1.4e+02;			
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;			
Qy 1 KKRIMHC 7			
Db 236 KRRLHLC 242			
: :			
RESULT 5			
AAG48145			
ID AAG48145 standard; Protein; 533 AA.			
XX AC AAG48145;			
XX DT 18-OCT-2000 (first entry)			
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60769.			
XX KW Protein identification; signal transduction pathway; metabolic pathway;			
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;			
XX KW termination sequence.			
XX OS Arabidopsis thaliana.			
XX PN EP1033405-A2.			
XX PD 06-SEP-2000.			
XX PF 25-FEB-2000; 2000EP-0301439.			
XX PR 25-FEB-1999; 99US-0121825.			
PR 05-MAR-1999; 99US-0123180.			
PR 09-MAR-1999; 99US-0123548.			
PR 23-MAR-1999; 99US-0125788.			
PR 25-MAR-1999; 99US-0126264.			
PR 29-MAR-1999; 99US-0126785.			
PR 01-APR-1999; 99US-0127462.			
PR 06-APR-1999; 99US-0128234.			
PR 08-APR-1999; 99US-0128714.			
PR 16-APR-1999; 99US-0129845.			
PR 19-APR-1999; 99US-0130077.			

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

```
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160840.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      80.5%; Score 33; DB 21; Length 533;
Best Local Similarity 57.1%; Pred. No. 1.5e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
Db 261 KRLLHC 267

RESULT 6
AAG48144
ID AAG48144 standard; Protein: 564 AA.
XX
AC AAG48144;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60768.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
```

PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.

PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 564;
Best Local Similarity 57.1%; Pred. No. 1.6e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
| : : : : |
Db 292 KKRLLHC 298

RESULT 7

AAR93552
ID AAR93552 standard; Peptide; 40 AA.

XX AAR93552;

XX AC AAR93552;
XX DT 25-JUN-1996 (first entry)

XX DE Bovine FKBP-13 immunophilin N-terminal peptide.
XX KW FKBP-13; immunophilin; FK506; rapamycin; rheumatoid arthritis;

XX KW diabetes; organ transplant; graft versus host disease;
XX KW immunosuppressant.
XX OS Bos taurus.

XX XX US5498597-A.

XX XX 12-MAR-1996.

XX PF 17-JAN-1992; 92US-0822966.

XX PR 17-JAN-1992; 92US-0822966.

XX XX (DAND) DANA FARBER CANCER INST INC.

XX PA (HARD) HARVARD COLLEGE.

XX PI Bierer BE, Burakoff SJ, Schreiber SL;

XX DR WPI; 1996-159713/16.

XX PT Purified mammalian FKBP-13 polypeptide capable of binding FK506
XX PT useful for identifying and studying immunosuppressant drugs

XX PS Disclosure; Column 8; 12pp; English.

XX CC This sequence encoding the bovine FKBP-13 N-terminal sequence
XX CC corresponds to the N-terminal sequence of human FKBP-13. FKBP-13
XX CC may be used for identifying immunosuppressant drugs, and may be
XX CC used in combination with immunosuppressant drugs for therapeutic
XX CC purposes in the treatment of autoimmune diseases e.g. rheumatoid
XX CC arthritis and type-I diabetes, organ transplant and graft versus
XX CC host disease. The recombinant form of the protein could be
XX CC potentially smaller and therefore easier to introduce into cells
XX CC than intact FKBP-13.

XX SQ Sequence 40 AA;

Query Match 78.0%; Score 32; DB 17; Length 40;
Best Local Similarity 71.4%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
DB 14 KKRVDHC 20

RESULT 8
ID ABG14241 standard; Protein: 83 AA.
XX
AC ABG14241;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #14232.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
WPI; 2001-639362/73.
DR N-PSDB; AAS78428.
XX
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
XX
Claim 20; SEQ ID No 44600; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

QY Sequence 83 AA;
DB 14 KKRVDHC 20

Query Match 78.0%; Score 32; DB 22; Length 83;
Best Local Similarity 71.4%; Pred. No. 40;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Best Local Similarity 83.3%; Pred. No. 40;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKRIMHC 7
DB 38 KRIVHC 43

RESULT 9
ID AAG76114 standard; Protein: 83 AA.
XX
AC AAG76114;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6878.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 11.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
WPI; 2001-235357/24.
DR N-PSDB; AAH35519.
XX
Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
useful for preventing, diagnosing and/or treating colorectal cancers -
XX
Claim 11; Page 8320-8322; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
cancer-associated nucleic acid molecules (N) and proteins (P), where
the proteins are collectively known as colon cancer antigens. The colon
cancer antigens have cytostatic activity and can be used in gene
therapy and vaccine production. N and P may be used in the prevention,
diagnosis and treatment of diseases associated with inappropriate P
expression. For example, N and P may be used to treat disorders
associated with decreased expression by rectifying mutations or deletions
in a patient's genome that affect the activity of P by expressing
inactive proteins or to supplement the patients own production of P.
Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

QY Sequence 83 AA;
DB 48 KKRVDHC 54

Query Match 78.0%; Score 32; DB 22; Length 83;
Best Local Similarity 71.4%; Pred. No. 40;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 10
ID AAY97245 standard; Protein; 94 AA.
XX
AC AAY97245;
XX
DT 19-DEC-2000 (first entry)
XX
DE Human INGB1 isoform N-terminal peptide.
XX
KW p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
KW INGI1; ing1; p33ING1; p37ING1; oncogene; gene therapy; diagnosis;
KW proliferation disorder; transformation; transformed cell; mouse.
XX
OS Homo sapiens.
XX
PN WO200046370-A1.
XX
PD 10-AUG-2000.
XX
PF 04-FEB-2000; 2000WO-US02959.
XX
PR 04-FEB-1999; 99US-0118941.
XX
PA (UNII) UNIV ILLINOIS FOUND.
XX
PI Gudkov A, Zeremski M, Gurova KV, Grigorian IA;
XX
DR WPI; 2000-491278/43.
XX
PT Detecting nucleic acid encoding exon 1b of ing1, useful for diagnosing
PT and treating cancer, comprises contacting sample with isolated nucleic
PT acid comprising sequence of exon 1b and detecting hybridized products
XX
PS Claim 26; Figure 7a; 134pp; English.
XX
CC Mutations in or loss of the p53 gene occur in more than 50% of
CC human tumours and tumour cell lines, but functional inactivation of
CC the p53 pathway occurs in a much larger proportion of tumours. In
CC many cases the mechanism of functional inactivation of the p53 gene
CC remains unknown but p53 has been found to act in cooperation with
CC INGI1. Functional cooperation between INGI1 and p53 suggested that
CC INGI1 encoded a tumour suppressor protein that functioned within the
CC p53 pathway. This data suggested a possible role for INGI1 in head
CC and neck cancers and chromosomal location of the INGI1 placed it
CC within a region that is frequently rearranged in head and neck
CC cancers. Large scale analysis of tumours involving INGI1 has not
CC revealed mutations in INGI1 nor significant variations in its
CC expression suggesting that INGI1 was not a useful gene to study in
CC cancer etiology. However, alternative initiation exons of the ing1
CC gene, each having their own promoter have been discovered.
CC Expression of one promoter (1a) produces a protein identical to
CC INGI1. Expression of a second promoter (1b) produces a protein having
CC an identical C-terminal fragment to INGI1 but an additional 104
CC N-terminal amino acids. The newly discovered protein has been
CC designated p37ING1 (Wild type: p33ING1). p37ING1 has the
CC characteristics of an oncogene. When overexpressed in cells (even
CC those expressing wild type p53) p37ING1 is able to cause
CC proliferation or transformation of those cells. Thus detecting a
CC nucleic acid encoding exon 1b of ing1 by hybridisation with an
CC isolated nucleic acid having the sequence of exon 1b of ing1
CC or its antisense sequence can identify individuals expressing the
CC oncogenic form of ing1. Novel peptide sequences taken from the 104
CC N-terminal peptide of p37ING1 can also be used to raise antibodies
CC that can also be used in detection methods for the p37ING1 variant.
CC The polypeptides may be useful in gene therapy for treatment of cell
CC proliferation disorders, especially cancers and for diagnosing and
CC studying cancers.
XX
SQ Sequence 94 AA;

Query Match 78.0%; Score 32; DB 21; Length 94;
Best Local Similarity 57.1%; Pred. No. 45;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKRIMHC 7
DB 67 KRRMLHC 73
RESULT 11
AAR28979
ID AAR28979 standard; protein; 99 AA.
XX
AC AAR28979;
XX
DT 24-MAR-1993 (first entry)
XX
DE Bovine RFBKP.
XX
KW Rapamycin; FK506; binding protein; RFBKP; prollyl isomerase;
KW immunosuppressant; cyclosporin A; macrolide; bovine; thymus; bRFBKP;
KW cis-trans prollyl isomerase activity; FKBP12.
XX
OS Bos taurus.
XX
PN WO9219745-A.
XX
PD 12-NOV-1992.
XX
PF 07-MAY-1992; 92WO-US03993.
XX
PR 08-MAY-1991; 91US-0697113.
XX
PA (VERT-) VERTEX PHARM INC.
XX
PI Harding MW;
XX
DR WPI; 1992-398871/48.
XX
PT New prollyl isomerase and rapamycin FK506 binding protein - useful
PT for screening potential immunosuppressive cpds.
XX
PS Disclosure; Fig 1; 30pp; English.
XX
CC This sequence corresponds to a fragment of a rapamycin FK506 binding
CC protein (RFBKP). RFBKP is a prollyl isomerase structurally related to
CC FK506 which does not bind the immunosuppressive cyclosporin A. RFBKP
CC binds FK506 and rapamycin with quantitatively significant selectivity.
CC RFBKP may be used in screening assays to detect new immunosuppressants
CC and to differentiate rapamycin-like cpds. from FK506-like cpds.
CC Rapamycin is a macrolide which is structurally related to FK506.
CC This RFBKP has been isolated from bovine thymus (bRFBKP) and was found
CC to be of low molecular weight, approx. 16,000, and to have cis-trans
CC prollyl isomerase activity. The N terminal of bRFBKP has been shown to
CC have over 50% homology to the N terminal of FKBP12.
XX
SQ Sequence 99 AA;
Query Match 78.0%; Score 32; DB 13; Length 99;
Best Local Similarity 71.4%; Pred. No. 47;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KKRIMHC 7
DB 14 KRRVDHC 20
RESULT 12
AAG03758
ID AAG03758 standard; Protein; 104 AA.
XX
AC AAG03758;
XX

DT XX 06-OCT-2000 (first entry)
 DE XX Human secreted protein, SEQ ID NO: 7839.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EPI033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 21-FEB-2000; 2000EP-0200610.
 PF 26-FEB-1999; 99US-0122487.
 XX
 PR (GEST) GENSET.
 XX
 PA Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 PI WPI: 2000-500381/45.
 XX
 DR N-PSDB: AAC03764.
 XX
 DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PT
 XX Claim 13; SEQ ID 7839; 71pp + CD-ROM; English.
 PS
 XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dr primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX
 SQ Sequence 104 AA;
 Query Match 78.0%; Score 32; DB 21; Length 104;
 Best Local Similarity 71.4%; Pred. No. 50;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 KKRIMHC 7
 Db 36 KKRVDHC 42
 III: II
 RESULT 13
 AAY03245
 ID AAY03245 standard; Protein; 128 AA.
 XX
 AC AAY03245;
 XX
 DT 26-AUG-1999 (first entry)
 XX
 DE Amino acid sequence of the human p28-ING1 protein.
 XX
 KW Human; p33-ING1 protein; growth regulation; apoptosis; DNA damage;
 KW inhibition; anchorage independent growth; cytotoxic drug; cancer;
 KW transcriptional activation; immortal cell line; p28-ING1 protein.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc_Difference 59...269

FT XX /note- "p36-ING1 fragment"
 PN W09916790-A1.
 XX
 PD 08-APR-1999.
 XX
 PF 24-SEP-1998; 98WO-US18179.
 XX
 XX 14-JAN-1998; 98US-0006783.
 PR 26-SEP-1997; 97US-0060138.
 XX
 XX (UNII) UNIV ILLINOIS BOARD OF TRUSTEES.
 PA (UYTE-) UNIV TECHNOLOGIES INT INC.
 XX
 XX Garkavtsev I, Gudkov A, Riabowol K;
 DR WPI: 1999-263685/22.
 XX
 PT Use of p33-ING1 peptides
 XX
 PS Example 8; Page 61; 64pp; English.
 CC
 CC This is the amino acid sequence of the human p28-ING1 protein,
 CC used in the method of the invention, involving the human p33-ING1
 CC protein. The ING1 gene encodes p33-ING1 which can be used to
 CC modulate the activity of, isolate or detect p53. Expression of the
 CC ING1 and p53 genes in a mammalian cell results in normal growth
 CC regulation anchorage-dependent growth and apoptosis as a response
 CC to irreversible DNA damage and other cellular insult. Inhibition of
 CC expression of either gene results in a loss of cellular growth
 CC control, anchorage independent growth, inhibition of apoptosis
 CC and resistance to radiation and cytotoxic drugs. The p33-ING1 is a
 CC component of the p53 signalling pathway that cooperates with p53 in
 CC negative regulation of cell proliferation by modulating p53 dependent
 CC transcriptional activation. Biological function of p53 signalling
 CC pathway can therefore be regulated (both enhanced or suppressed) by
 CC modulating p33-ING1 activity. The modulation of p33-ING1 activity can
 CC be used for the stimulation or restoration of the p53 pathway in
 CC anti cancer therapy or for the suppression of the p53 pathway in
 CC defend sensitive tissues from genotoxic stress or for the generation
 CC of immortal cell lines.
 XX
 SQ Sequence 128 AA;
 Query Match 78.0%; Score 32; DB 20; Length 128;
 Best Local Similarity 57.1%; Pred. No. 60;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 KKRIMHC 7
 Db 67 KKRMLHC 73
 I:::II
 RESULT 14
 AAB87656
 ID AAB87656 standard; protein; 137 AA.
 XX
 AC AAB87656;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Bovine mammary tissue derived protein #47.
 DE
 DE Bovine; mammary gland; cancer; tumour; angiogenesis.
 KW
 XX Bos taurus.
 OS
 XX W0200114553-A1.
 PN
 XX 01-MAR-2001.
 PD
 XX 23-AUG-2000; 2000WO-N200166.
 PF
 XX

PR 23-AUG-1999; 99US-0150330.
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
XX
XX Havukkala IJ, Gleen M, Grigor MR, Molenaar AJ;
XX WPI; 2001-226619/23.
XX
XX New polypeptides and polynucleotides encoding the polypeptides, which
PT are expressed in bovine mammary gland tissue, useful for stimulating
PT mammary gland growth or function, or inducing differentiation of milk
PT producing cells -
XX
XX Claim 11; Page 80; 97pp; English.
XX
XX The present invention relates to proteins derived from bovine
CC mammary gland cells. The invention is useful for stimulating
CC bovine mammary gland cell growth and function, inhibiting the
CC growth of various mammary gland cancer cells, inhibiting
CC angiogenesis and vascularization of tumours, or modulating
CC the growth of blood vessels in a mammal.
XX
SQ Sequence 137 AA;
Query Match 78.0%; Score 32; DB 22; Length 137;
Best Local Similarity 71.4%; Pred. No. 64;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KKRIMHC 7
DB 43 KKRVDHC 49
III: II
RESULT 15
AAR93551
ID AAR93551 standard; Protein: 141 AA.
XX
AC AAR93551;
XX
DT 25-JUN-1996 (first entry)
XX
XX Human FKBP-13 immunophilin protein.
XX
XX FKBP-13; immunophilin; FK506; rapamycin; rheumatoid arthritis;
KW diabetes; organ transplant; graft versus host disease;
KW immunosuppressant.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT Peptide 1..21
FT /label= sig_peptide
FT Misc-difference 22..61
FT /note= "corresponds to bovine thymus FKBP-13 N-
FT terminal sequence"
FT Misc-difference 117..120
FT /note= "endoplasmic reticulum retention sequence"
XX
PN US5498597-A.
XX
XX 12-MAR-1996.
XX
XX 17-JAN-1992; 92US-0822966.
XX
XX 17-JAN-1992; 92US-0822966.
XX
XX (DAND) DANA FARBER CANCER INST INC.
PA (HARD) HARVARD COLLEGE.
XX
PI Bierer BE, Burakoff SJ, Schreiber SL;
XX WPI; 1996-159713/16.
DR

DR N-PSDB; AAT18037.
XX
XX Purified mammalian FKBP-13 polypeptide capable of binding FK506
PT useful for identifying and studying immunosuppressant drugs
XX
XX Claim 1; Fig.1; 12pp; English.
XX
XX The FKBP-13 protein may be used for identifying immunosuppressant
CC drugs, and may be used in combination with immunosuppressant drugs
CC for therapeutic purposes in the treatment of autoimmune diseases e.g.
CC rheumatoid arthritis and type-1 diabetes, organ transplant and
CC graft versus host disease. The recombinant form of the protein
CC could be potentially smaller and therefore easier to introduce
CC into cells than intact FKBP-13.
XX
XX Sequence 141 AA;
SQ
Query Match 78.0%; Score 32; DB 17; Length 141;
Best Local Similarity 71.4%; Pred. No. 66;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KKRIMHC 7
DB 35 KKRVDHC 41
III: II
Search completed: November 1, 2002, 14:47:06
Job time : 3.58682 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:08 ; Search time 1.45338 Seconds
(without alignments)
952.238 Million cell updates/sec

Title: US-09-606-129A-19
Perfect score: 46
Sequence: 1 QKLCCHQKK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Result No.	Score	Query %	Length	ID	Description
1	37	80.4	636	12 O90142	O90142 spodoptera
2	37	80.4	1063	12 Q9J844	Q9J844 spodoptera
3	35	76.1	295	11 Q9CY64	Q9CY64 mus musculus
4	35	76.1	303	11 Q9DD21	Q9DD21 mus musculus
5	35	76.1	420	10 O22019	O22019 cyanidiosch
6	34	73.9	208	5 Q94164	Q94164 caenorhabdi
7	34	73.9	322	4 Q9BSM2	Q9BSM2 homo sapien
8	34	73.9	409	11 Q9D3R6	Q9D3R6 mus musculus
9	34	73.9	522	4 Q9H8F4	Q9H8F4 homo sapien
10	34	73.9	547	3 Q74308	Q74308 schizosacch
11	34	73.9	790	4 Q9NVK9	Q9NVK9 homo sapien
12	34	73.9	1188	5 Q96143	Q96143 plasmodium
13	34	73.9	1295	3 O13348	O13348 magnaporthe
14	34	73.9	1473	11 Q63625	Q63625 rattus norv
15	34	73.9	1654	4 Q9P1Y6	Q9P1Y6 homo sapien
16	33	71.7	261	4 Q9H1N6	Q9H1N6 homo sapien

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	37	80.4	636	12 O90142	O90142 spodoptera
2	37	80.4	1063	12 Q9J844	Q9J844 spodoptera
3	35	76.1	295	11 Q9CY64	Q9CY64 mus musculus
4	35	76.1	303	11 Q9DD21	Q9DD21 mus musculus
5	35	76.1	420	10 O22019	O22019 cyanidiosch
6	34	73.9	208	5 Q94164	Q94164 caenorhabdi
7	34	73.9	322	4 Q9BSM2	Q9BSM2 homo sapien
8	34	73.9	409	11 Q9D3R6	Q9D3R6 mus musculus
9	34	73.9	522	4 Q9H8F4	Q9H8F4 homo sapien
10	34	73.9	547	3 Q74308	Q74308 schizosacch
11	34	73.9	790	4 Q9NVK9	Q9NVK9 homo sapien
12	34	73.9	1188	5 Q96143	Q96143 plasmodium
13	34	73.9	1295	3 O13348	O13348 magnaporthe
14	34	73.9	1473	11 Q63625	Q63625 rattus norv
15	34	73.9	1654	4 Q9P1Y6	Q9P1Y6 homo sapien
16	33	71.7	261	4 Q9H1N6	Q9H1N6 homo sapien

Query Match 80.4%; Score 37; DB 12; Length 636;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 2 KLCCHQKK 8
|||||:

17	33	71.7	263	17 Q9YDU2	Q9YDU2 aeropyrum p
18	33	71.7	266	2 P96301	P96301 alcaligenes
19	33	71.7	273	12 Q9PYW7	Q9PYW7 xestia c-ni
20	33	71.7	445	10 Q9ZPL5	Q9ZPL5 nicotiana t
21	33	71.7	540	8 Q9MT29	Q9MT29 solanum tub
22	33	71.7	566	4 Q9H2S0	Q9H2S0 homo sapien
23	33	71.7	588	4 Q9ULU1	Q9ULU1 homo sapien
24	33	71.7	673	11 Q9QVT4	Q9QVT4 mus sp. mpl
25	33	71.7	773	10 Q80631	Q80631 arabidopsis
26	33	71.7	1247	5 Q17461	Q17461 caenorhabdi
27	32	69.6	52	16 Q98C26	Q98C26 rhizobium 1
28	32	69.6	87	5 Q9NSN6	Q9NSN6 caenorhabdi
29	32	69.6	125	4 Q96KV9	Q96KV9 homo sapien
30	32	69.6	159	4 Q96KW0	Q96KW0 homo sapien
31	32	69.6	163	2 Q9EUC2	Q9EUC2 escherichia
32	32	69.6	221	17 Q97ZS9	Q97ZS9 sulfolobus
33	32	69.6	229	17 Q97GH5	Q97GH5 sulfolobus
34	32	69.6	312	11 Q9CSE2	Q9CSE2 mus musculu
35	32	69.6	336	4 Q9H495	Q9H495 homo sapien
36	32	69.6	380	5 Q21866	Q21866 caenorhabdi
37	32	69.6	397	4 Q9H497	Q9H497 homo sapien
38	32	69.6	397	4 Q9H6E7	Q9H6E7 homo sapien
39	32	69.6	422	10 Q9SSP1	Q9SSP1 arabidopsis
40	32	69.6	462	5 Q9U0E3	Q9U0E3 strongyloce
41	32	69.6	465	5 Q9XW94	Q9XW94 caenorhabdi
42	32	69.6	474	10 Q80588	Q80588 arabidopsis
43	32	69.6	544	10 Q943X8	Q943X8 oryza sativ
44	32	69.6	548	5 Q20367	Q20367 caenorhabdi
45	32	69.6	590	2 Q45490	Q45490 bacillus st

ALIGNMENTS

RESULT 1					
O90142	ID	O90142	PRELIMINARY;	PRT;	636 AA.
AC	O90142;				
DT	O1-NOV-1998	(TREMBLrel. 08, Created)			
DT	O1-NOV-1998	(TREMBLrel. 08, Last sequence update)			
DT	O1-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	DNA POLYMERASE (FRAGMENT).				
GN	DPOL.				
OS	Spodoptera exigua nucleopolyhedrovirus.				
OC	Viruses; dsDNA viruses, no RNA stage; Baculoviridae;				
OC	Nucleopolyhedrovirus.				
OX	NCBI_TaxID=10454;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Bulach D.M., Kumar C.A., Zaia A., Liang B., Tribe D.E.;				
RT	"Group II Nucleopolyhedrovirus Subgroups Revealed by Phylogenetic				
RT	Analysis of Polyhedrin and DNA Polymerase Gene Sequences.;"				
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.				
CC	-!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N				
CC	PYROPHOSPHATE + DNA(N).				
CC	-!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.				
DR	EMBL; AF068186; AAC33749.1; -				
DR	InterPro; IPR002064; DNA_pol_B.				
DR	Pfam; PF00136; DNA_pol_B; 1.				
DR	Pfam; PF03104; DNA_pol_B-exo; 1.				
DR	PRINTS; PR00106; DNAPOLB.				
DR	SMART; SM00486; POLBc; 1.				
KW	DNA replication; DNA-binding; DNA-directed DNA polymerase.				
FT	NON_TER	1			
FT	NON_TER	636			
SQ	SEQUENCE	636 AA;	73837 MW;	9BDB36475B07DCD1	CRC64;

Query Match 80.4%; Score 37; DB 12; Length 636;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 2 KLCCHQKK 8
|||||:

```

Db 392 KLCHQK 398
RESULT 2
Q9J844 PRELIMINARY; PRT; 1063 AA.
AC Q9J844;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF93 DNA POLYMERASE.
OS Spodoptera exigua nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=10454;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20036646; PubMed=10567663;
RX IJkel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,
RA Goldbach R.W., Vlask J.M.;
RT "Sequence and organization of the spodoptera exigua multicapsid
RT nucleopolyhedrovirus genome.";
RL J. Gen. Virol. 80:3289-3304(1999).
[2]
RN SEQUENCE FROM N.A.
RP IJkel W.F.J., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,
RA Goldbach R.W., Vlask J.M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N
CC PYROPHOSPHATE + DNA(N).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC EMBL; AF169823; AAF33622.1; -
DR InterPro: IPR002064; DNA_pol_B.
DR Pfam; PF031136; DNA_pol_B; 1.
DR Pfam; PF031104; DNA_pol_B_exo; 1.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLBc; 1.
DR DNA replication; DNA-binding; DNA-directed DNA polymerase.
KW SEQUENCE 1063 AA; 123109 MW; 471603FAA92B9A10 CRC64;
SQ SEQUENCE 1063 AA; 123109 MW; 471603FAA92B9A10 CRC64;

Query Match 80.4%; Score 37; DB 12; Length 1063;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLCHQK 8
|:|:|:|:|
Db 557 KLCHQK 563

RESULT 3
Q9CY64 PRELIMINARY; PRT; 295 AA.
AC Q9CY64;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 2500001N03RIK PROTEIN.
GN 2500001N03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002231; BAB21950.1; -
DR MGD; MGI:1915580; 0610006Allirik.
DR InterPro: IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
SQ SEQUENCE 303 AA; 34491 MW; 52D8A3B02EE956EB CRC64;

Query Match 76.1%; Score 35; DB 11; Length 303;
Best Local Similarity 62.5%; Pred. No. 25;

```

```

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010847; BAB27219.1; -
DR MGD; MGI:1917355; 2500001N03RIK.
DR InterPro: IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
SQ SEQUENCE 295 AA; 33524 MW; F2E1682BD77032A4 CRC64;

Query Match 76.1%; Score 35; DB 11; Length 295;
Best Local Similarity 62.5%; Pred. No. 24;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCQK 8
|:|:|:|:|
Db 288 QKLCQK 295

RESULT 4
Q9DD21 PRELIMINARY; PRT; 303 AA.
AC Q9DD21;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 0610006Allirik PROTEIN.
GN 0610006Allirik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002231; BAB21950.1; -
DR MGD; MGI:1915580; 0610006Allirik.
DR InterPro: IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
SQ SEQUENCE 303 AA; 34491 MW; 52D8A3B02EE956EB CRC64;

Query Match 76.1%; Score 35; DB 11; Length 303;
Best Local Similarity 62.5%; Pred. No. 25;

```

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCHKK 8

Db 288 QRLCHRKQ 295

RESULT 5

O22019
ID O22019 PRELIMINARY; PRT; 420 AA.
AC O22019; 1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF420.
OS Cyanidioschyzon merolae.
OC Eukaryota; Rhodophyta; Bangiophyceae; Cyanidioschyzon.
OX NCBI_TaxID=45157;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohta N.;
RL J. Plant Res. 110:235-245(1997).
DR EMBL; D63675; BAA22815.1; -.
DR InterPro; IPR004161; GTP_EFTU_D2.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR001950; SUIL.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR ProDom; PD186100; IF2; 1.
DR PROSITE; PS01118; SUIL_1; UNKNOWN_1.
SQ SEQUENCE 420 AA; 47691 MW; A6CAE107B2484E19 CRC64;

Query Match 76.1%; Score 35; DB 10; Length 420;

Best Local Similarity 85.7%; Pred. No. 34;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHKK 7

Db 236 QKLCHKP 242

RESULT 6

O94164
ID O94164 PRELIMINARY; PRT; 208 AA.
AC O94164;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 24.0 KDA PROTEIN.
GN C10G8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Blanchard M., Bradshaw H.;
RL Investigating biology. The C. elegans cosmid C10G8.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Blanchard M., Bradshaw H.;
RL "The sequence of C. elegans cosmid C10G8.";
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

*Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
Science 282:2012-2018(1998).

DR EMBL; U70857; AAB09170.1; -.
DR InterPro; IPR002223; Kunitz_BPTI.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR SMART; SM00131; KU; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 1.
KW Hypothetical protein; Serine protease inhibitor.
SQ SEQUENCE 208 AA; 24008 MW; 779AB4A8948E67B0 CRC64;

Query Match 73.9%; Score 34; DB 5; Length 208;

Best Local Similarity 83.3%; Pred. No. 28;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LCHQKK 8

Db 172 MCHQKK 177

RESULT 7

O9BSM2
ID O9BSM2 PRELIMINARY; PRT; 322 AA.
AC O9BSM2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 34.9 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LYMPH, LYMPHOMA;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004950; AAH04950.1; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 322 AA; 34916 MW; 33C65E86B8F9D761 CRC64;

Query Match 73.9%; Score 34; DB 4; Length 322;

Best Local Similarity 71.4%; Pred. No. 42;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHKK 7

Db 274 QKLCHKK 280

RESULT 8

O9D3R6
ID O9D3R6 PRELIMINARY; PRT; 409 AA.
AC O9D3R6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 4933439B08RIK PROTEIN.
GN 4933439B08RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Izawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

```

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK017114; BAB30604.1; -.
DR MGD; MGI:1918456; 4933439B08Rik.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003959; AAA_subfam.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
KW ATP-binding.
SQ SEQUENCE 409 AA; 46131 MW; 9600B20008DC9749 CRC64;

Query Match 73.9%; Score 34; DB 11; Length 409;
Best Local Similarity 83.3%; Pred. No. 52;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLCHKQK 7
Db 122 KICHQK 127
:||||:

RESULT 9
Q9H8F4 PRELIMINARY; PRT; 522 AA.
AC Q9H8F4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CDNA FLJ13679 FIS, CLONE PLACE2000006.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RA "NEDO human cDNA sequencing project.";
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK023741; BAB14663.1; -.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhGEF.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhGEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhGEF; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
SQ SEQUENCE 522 AA; 59222 MW; F01A6BF70D2920F9 CRC64;

Query Match 73.9%; Score 34; DB 4; Length 522;
Best Local Similarity 62.5%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCHKQK 8
Db 157 QELCHQOE 164
:||||:

RESULT 10

```

```

074308 PRELIMINARY; PRT; 547 AA.
ID 074308;
AC 074308;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL ZINC-FINGER PROTEIN.
GN SPBC15D4.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC NCBI_TaxID=4896;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;
RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RL -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE BINUCLEAR CLUSTER DOMAIN.
DR EMBL; AL031349; CAA20477.1; -.
DR HSSP; P25502; IZME.
DR InterPro; IPR001138; Zn2_Cy6_fungal.
DR Pfam; PF00172; Zn_Clus; 1.
DR SMART; SM00066; GAL4; 1.
DR PROSITE; PS00463; ZN2_Cy6_FUNGAL_1; 1.
DR PROSITE; PS50048; ZN2_Cy6_FUNGAL_2; 1.
KW DNA-binding; Metal-binding; Nuclear protein; Transcription regulation;
KW Zinc.
SQ SEQUENCE 547 AA; 59641 MW; A65FD7D03996CD18 CRC64;

Query Match 73.9%; Score 34; DB 3; Length 547;
Best Local Similarity 71.4%; Pred. No. 67;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCHKQK 7
Db 427 RKLCHK 433
:||||:

RESULT 11
Q9NVK9 PRELIMINARY; PRT; 790 AA.
AC Q9NVK9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CDNA FLJ10665 FIS, CLONE NT2RP2006200.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA Masuho Y., Kanehori K.;
RA "NEDO human cDNA sequencing project.";
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK001527; BAA91741.1; -.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhGEF.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhGEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhGEF; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
SQ SEQUENCE 790 AA; 88976 MW; A3B9F4972CE2D509 CRC64;

Query Match 73.9%; Score 34; DB 4; Length 790;
Best Local Similarity 62.5%; Pred. No. 94;

```

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCQKK 8
:|||||
Db 157 QELCHQKE 164

RESULT 12

O96143 ID O96143 PRELIMINARY; PRT; 1188 AA.
AC O96143;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PROTEIN WITH 5'-3' EXONUCLEASE DOMAIN (KEM-1 FAMILY).
GN PRB0205C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL; AE001380; AAC71830.1; -.
KW Exonuclease.

SQ SEQUENCE 1188 AA; 142895 MW; BF767FC8532EBAC9 CRC64;

Query Match 73.9%; Score 34; DB 5; Length 1188;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCQKK 7
:|||||
Db 1117 QELCHQK 1123

RESULT 13

O13348 ID O13348 PRELIMINARY; PRT; 1295 AA.
AC O13348;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE REVERSE TRANSCRIPTASE
OS Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
OX NCBI_TaxID=148305;
RN [1]
RP SEQUENCE FROM N.A.
RA Meyn III M.A., Farrall L., Valent B., Chumley F.G., Orbach M.J.;
RT "Magnaporthe grisea repeated DNA element MGR583 is a member of the
LINE-1 class of polyA retrotransposons.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF018033; AAB71689.1; -.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1295 AA; 144308 MW; E811059B750D5421 CRC64;

Query Match 73.9%; Score 34; DB 3; Length 1295;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KLCHQKK 8
:|||||
Db 33 ELCHQKK 39

RESULT 14

O63625 ID O63625 PRELIMINARY; PRT; 1473 AA.
AC O63625;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RA9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=HIPPOCAMPUS;
RX MEDLINE=96293459; PubMed=8692929;
RA Yuryev A., Patturajan M., Litington Y., Joshi R.V., Gentile C.,
RA Gebara M., Corden J.L.;
RT "The C-terminal domain of the largest subunit of RNA polymerase II
interacts with a novel set of serine/arginine-rich proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:6975-6980(1996).
DR EMBL; U49057; AAC52658.1; -.
SQ SEQUENCE 1473 AA; 161204 MW; 949EE6F5873989BF CRC64;

Query Match 73.9%; Score 34; DB 11; Length 1473;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCQKK 7
:|||||
Db 1425 QKICHSK 1431

RESULT 15

O9PIY6 ID O9PIY6 PRELIMINARY; PRT; 1654 AA.
AC O9PIY6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1542 PROTEIN (FRAGMENT).
GN KIAA1542.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
genes.XVII.The complete sequences of 100 new cDNA clones from brain
DNA Res. 7:143-150(2000).
DR EMBL; AB040975; BAA96066.1; -.
DR HSSP; P28990; 1CHC.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00109; C1; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00184; RING; 2.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
KW Zinc-finger.
FT NON_TER 1
SQ SEQUENCE 1654 AA; 179053 MW; 125CF71A84AFB218 CRC64;

Query Match 73.9%; Score 34; DB 4; Length 1654;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHQK 7
||:| |
Db 1606 QKCHSK 1612

Search completed: November 1, 2002, 14:49:49
Job time : 4.45338 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:41:32 ; Search time 0.411576 Seconds
(without alignments)
752.611 Million cell updates/sec

Title: US-09-606-129A-19
Perfect score: 46
Sequence: 1 QKLCHQKK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	46	100.0	295	1 BIEA_RAT	P46844 rattus norv
2	36	78.3	654	1 CBPI_YEAST	P07252 saccharomyc
3	33	71.7	673	1 Z145_HUMAN	Q05516 homo sapien
4	32	69.6	135	1 VEG_PAPVE	P11331 european el
5	32	69.6	210	1 YOS3_CAEEL	Q09311 caenorhabdi
6	32	69.6	272	1 3DHQ_ACICA	Q59087 acinetobact
7	32	69.6	272	1 DHPF_CLOPE	Q46185 clostridium
8	32	69.6	345	1 TF2B_YEAST	P29055 saccharomyc
9	32	69.6	357	1 Y303_MYCGE	P47545 mycoplasma
10	32	69.6	595	1 PRIM_CHLTR	O84799 chlamydia t
11	32	69.6	609	1 FETA_GORGO	P28050 gorilla gor
12	32	69.6	609	1 FETA_HUMAN	P02771 homo sapien
13	32	69.6	645	1 Y081_CAEEL	P34617 caenorhabdi
14	32	69.6	895	1 SY1_MYCGE	P47587 mycoplasma
15	32	69.6	1593	1 AT12_HUMAN	P58397 homo sapien
16	31	67.4	75	1 EX7S_BACSU	P54522 bacillus su
17	31	67.4	169	1 NEUT_BOVIN	P01156 bos taurus
18	31	67.4	170	1 NEUT_CANFA	P10673 canis fami
19	31	67.4	271	1 YSM4_CAEEL	Q10124 caenorhabdi
20	31	67.4	478	1 ARDE_CHLTR	P47351 chlamydia t
21	31	67.4	498	1 NFS1_SCHPO	O74351 schizosacch
22	31	67.4	631	1 Z179_RAT	O70418 rattus norv
23	31	67.4	677	1 NRGL_XENLA	Q93383 xenopus lae
24	31	67.4	892	1 L0L2_XENLA	Q93088 xenopus lae
25	31	67.4	1207	1 DPOL_ASFB7	P42489 african swi
26	31	67.4	1244	1 DPOL_ASFL6	P43139 african swi
27	31	67.4	1256	1 FLII_DROME	Q24020 drosophila
28	31	67.4	1976	1 MYHA_BOVIN	Q27991 bos taurus
29	31	67.4	1976	1 MYHA_HUMAN	P35580 homo sapien
30	31	67.4	1976	1 MYHA_RAT	Q9310 rattus norv
31	30	65.2	80	1 EX7S_BACHD	Q9K968 bacillus ha
32	30	65.2	159	1 MPAA_CORAV	Q08407 corylus ave
33	30	65.2	259	1 DEOC_ECOLI	P00882 escherichia

ALIGNMENTS

RESULT 1

ID	BIEA_RAT	STANDARD;	PRT;	295 AA.
AC	P46844;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Biliverdin reductase A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-reductase).			
DE	reductase).			
GN	BLVRA OR BLVR.			
OS	Rattus norvegicus (Rat).			
OC	Eumariota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Kidney;			
RX	MEDLINE=92156147; PubMed=1371282;			
RT	Fakhrai H., Maines M.D.;			
RT	"Expression and characterization of a cDNA for rat kidney biliverdin reductase. Evidence suggesting the liver and kidney enzymes are the same transcript product."			
RT	same transcript product."			
RL	J. Biol. Chem. 267:4023-4029(1992).			
RN	[2]			
RP	MUTAGENESIS.			
RA	MEDLINE=94291657; PubMed=8020496;			
RA	McCoubrey W.K. Jr., Maines M.D.;			
RT	"Site-directed mutagenesis of cysteine residues in biliverdin reductase. Roles in substrate and cofactor binding."			
RL	Eur. J. Biochem. 222:597-603(1994).			
CC	- - FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN; DISPLAYS TWO DISTINCT PH OPTIMA USING A DIFFERENT COFACTOR AT EACH PH: NADH AT THE LOWER PH 6.7-6.9 RANGE AND NADPH AT PH 8.5-8.7. NADPH, HOWEVER, IS THE PROBABLE COFACTOR IN BIOLOGICAL SYSTEMS.			
CC	- - CATALYTIC ACTIVITY: Bilirubin + NAD(P)(+)- = biliverdin + NAD(P)H.			
CC	- - COFACTOR: BINDS ONE ZINC ION.			
CC	- - PATHWAY: FINAL STEP IN HEME METABOLISM.			
CC	- - SUBUNIT: MONOMER (BY SIMILARITY).			
CC	- - SUBCELLULAR LOCATION: Cytoplasmic.			
CC	- - SIMILARITY: TO E.COLI YHH.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M81681; AAA40830.1; -			
DR	InterPro: IPR000683; GFO_IDH_MocA.			
DR	Pfam: PF01408; GFO_IDH_MocA; 1.			
KW	Oxidoreductase; NAD; NADP; Zinc.			
PROPEP	1 2			
FT	CHAIN 1 3 295			
FT	DOMAIN 11 16			
FT	BILIVERDIN REDUCTASE A.			
FT	POLY-VAL.			

P75194 mycoplasma
P23352 visna lenti
P23424 visna lenti
P23425 visna lenti
P35955 visna lenti
P16900 ovine lenti
Q01767 streptomyces
P19580 bacillus an
Q26793 potocous tr
O57428 xenopus lae
Q10715 haematobia
P56410 anas platyr

34 30 65.2 347 1 YF86_MYCPN
35 30 65.2 442 1 GAG_VILV
36 30 65.2 442 1 GAG_VILV1
37 30 65.2 442 1 GAG_VILV2
38 30 65.2 442 1 GAG_VILVK
39 30 65.2 446 1 GAG_OMVVS
40 30 65.2 457 1 LAT_STRCL
41 30 65.2 464 1 CAPB_BACAN
42 30 65.2 489 1 OCLN_POTTR
43 30 65.2 583 1 HASS_XENLA
44 30 65.2 611 1 ACE_HAEIE
45 30 65.2 686 1 TRFE_ANAPL

FT METAL 279 ZINC (POTENTIAL).
FT METAL 280 ZINC (POTENTIAL).
FT METAL 291 ZINC (POTENTIAL).
FT METAL 292 ZINC (POTENTIAL).
FT METAL 293 ZINC (POTENTIAL).
FT MUTAGEN 73 C->A: LOSS OF ACTIVITY.
FT MUTAGEN 280 C->A: REDUCED ACTIVITY.
FT MUTAGEN 291 C->A: REDUCED ACTIVITY.
SQ SEQUENCE 295 AA; 33565 MW; 219C8EA96C150588 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. NO. 0.19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCHQKK 8

Db 288 QKLCHQKK 295

RESULT 2

CBP1_YEAST STANDARD; PRT; 654 AA.
AC P07252;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cytochrome B pre-mRNA processing protein 1.
GN CBP1 OR XTL209W OR J0242 OR HRA654.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84185566; PubMed=6325407;
RA Dieckmann C.B., Homslon G., Tzagoloff A.;
RT "Assembly of the mitochondrial membrane system. Nucleotide sequence of a yeast nuclear gene (CBP1) involved in 5' end processing of cytochrome b pre-mRNA."
RL J. Biol. Chem. 259:4732-4738(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
MEDLINE=95242842; PubMed=7725802;
RA Vandenbol M., Durand P., Bolle P.-A., Dion C., Portetelle D., Hilger F.;
RT "Sequence analysis of a 40.2 kb DNA fragment located near the left telomere of yeast chromosome X."
RL Yeast 10:1657-1662(1994).
RN [3]
RP SEQUENCE OF 589-654 FROM N.A.
RX MEDLINE=90014786; PubMed=2552292;
RA Liu Y., Dieckmann C.L.;
RT "Overproduction of yeast viruslike particles by strains deficient in a mitochondrial nuclease."
RL Mol. Cell. Biol. 9:3323-3331(1989).
CC -1- FUNCTION: RESPONSIBLE FOR CONFERRING A STABLE 5' END ON CYTOCHROME B MRNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; K02647; AAA34474.1; -
DR EMBL; Z34098; CAA84002.1; -
DR EMBL; Z49484; CAA89506.1; -
DR EMBL; M28067; AAA34456.1; -
DR PIR; S05829; BVDYPI.
DR PIR; S45164; S45164.
DR SGD; S0003745; CBP1.

KW mRNA processing.
SQ SEQUENCE 654 AA; 76171 MW; 2453B03280E1C44D CRC64;

Query Match 78.38; Score 36; DB 1; Length 654;
Best Local Similarity 75.0%; Pred. NO. 23;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHQKK 8

Db 514 KKLCHYKK 521

RESULT 3

Z145_HUMAN STANDARD; PRT; 673 AA.
AC Q05516;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein PLZF (Promyelocytic leukemia zinc finger protein) (Zinc finger protein 145).
GN ZNF145 OR PLZF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RX MEDLINE=93209216; PubMed=8384553;
RA Chen Z., Brand N.J., Chen A., Chen S.-J., Tong J.-H., Wang Z.-Y., Waxman S., Zelent A.;
RT "Fusion between a novel Kruppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) RT translocation associated with acute promyelocytic leukaemia.";
RL EMBO J. 12:1161-1167(1993).
RN [2]
RP SEQUENCE OF 424-455 FROM N.A.
RX MEDLINE=93253074; PubMed=8387545;
RA Chen S.-J., Zelent A., Tong J.-H., Yu H.-Q., Wang Z.-Y., Derre J., Berger R., Waxman S., Chen Z.;
RT "Rearrangements of the retinoic acid receptor alpha and promyelocytic leukemia zinc finger genes resulting from t(11;17)(q23;q21) in a patient with acute promyelocytic leukemia.";
RL J. Clin. Invest. 91:2260-2267(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 7-122.
RX MEDLINE=20005701; PubMed=10537309;
RA Li X., Peng H., Schultz D.C., Lopez-Guisa J.M., Rauscher F.J. III, Marmorstein R.;
RT "Structure-function studies of the BTB/POZ transcriptional repression domain from the promyelocytic leukemia zinc finger oncoprotein.";
RL Cancer Res. 59:5275-5282(1999).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. MAY PLAY A ROLE IN MYELOID MATURATION AND IN THE DEVELOPMENT AND/OR MAINTENANCE OF OTHER DIFFERENTIATED TISSUES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PLZF1 AND PLZF2 (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WITHIN THE HEMATOPOIETIC SYSTEM, PLZF IS EXPRESSED IN BONE MARROW, EARLY MYELOID CELL LINES AND PERIPHERAL BLOOD MONONUCLEAR CELLS. ALSO EXPRESSED IN THE OVARY, AND AT LOWER LEVELS, IN THE KIDNEY AND LUNG.
CC -1- INDUCTION: BY RETINOIC ACID.
CC -1- DISEASE: A FORM OF ACUTE PROMYELOCYTIC LEUKEMIA (APL) IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(11;17)(Q32;Q21) WHICH INVOLVES ZNF145 AND RETINOIC ACID RECEPTOR ALPHA (RARA).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).


```

RESULT 6
3DHQ_ACICA          STANDARD;          PRT;    272 AA.
ID 3DHQ_ACICA
AC Q9087;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Catabolic 3-dehydroquininate dehydratase (EC 4.2.1.10) (3-
DE dehydroquinase).
GN QIIB.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=BD413 / ADP1;
RX MEDLINE=96011389; PubMed=7592351;
RA Elsemore D.A., Ornstun L.N.;
RT "Unusual ancestry of dehydratases associated with quinate catabolism
RT in Acinetobacter calcoaceticus";
RL J. Bacteriol. 177:5971-5978(1995).
CC -1- FUNCTION: CATALYZES THE CATABOLIC DEHYDRATATION OF DEHYDROQUINATE
CC TO DEHYDROSHIKIMATE.
CC -1- CATALYTIC ACTIVITY: 3-dehydroquininate = 3-dehydroshikimate + H(2)O.
CC -1- PATHWAY: QUINIC ACID CATABOLIC PATHWAY; SECOND STEP. THIS PATHWAY
CC ALLOWS GROWTH OF BACTERIA WITH QUINATE BY ITS CONVERSION TO
CC PROTOCATECHUATE AND SUBSEQUENT METABOLISM BY THE BETA-KETOADIPATE
CC PATHWAY.
CC -1- INDUCTION: BY PROTOCATECHUATE.
CC -1- SIMILARITY: BELONGS TO THE TYPE-I 3-DEHYDROQUINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L05770; AAC37158.1; ALT_INIT.
CC InterPro; IPR001381; DHQuinase_I.
CC Pfam; PF01487; DHQuinase_I; 1.
CC PROSITE; PS01028; DEHYDROQUINASE_I; 1.
CC Quinate metabolism; Lyase.
CC ACT_SITE 163 163 BY SIMILARITY.
CC ACT_SITE 190 190 FORMS A SCHIFF-BASE INTERMEDIATE
CC (BY SIMILARITY).
CC -----
CC SEQUENCE 272 AA; 29899 MW; 98646DC5E88BF6D3 CRC64;
Query Match 69.6%; Score 32; DB 1; Length 272;
Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 KLCHQK 8
Db 149 KLAHQK 155
|||
|||

RESULT 7
DAPF_CLOPE          STANDARD;          PRT;    272 AA.
ID DAPF_CLOPE
AC Q46185;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Diaminopimelate epimerase (EC 5.1.1.7) (DAP epimerase).
GN DAPF OR CPE1846.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE OF 168-238 FROM N.A.
RC STRAIN=CPN50;
RX PubMed=7559358;
RA Katayama S.-I., Dupuy B., Garnier T., Cole S.T.;
RT "Rapid expansion of the physical and genetic map of the chromosome of
RT Clostridium perfringens CPN50";
RL J. Bacteriol. 177:5680-5685(1995).
CC -1- CATALYTIC ACTIVITY: LL-2,6-diaminoheptanedioate = meso-
CC diaminoheptanedioate.
CC -1- PATHWAY: SIXTH STEP IN THE BIOSYNTHESIS OF LYSINE FROM ASPARTATE
CC SEMIALDEHYDE.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DIAMINOPIMELATE EPIMERASE FAMILY.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 181.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP003192; BAB81552.1; -; ALT_FRAME.
CC EMBL; X86511; CAA60229.1; ALT_FRAME.
CC HSSP; P44859; 1BW2.
CC InterPro; IPR001653; DAP_Epimerase.
CC PROSITE; PS01326; DAP_EPIMERASE; 1.
CC Isomerase; Lysine biosynthesis; Complete proteome.
CC ACT_SITE 72 72 BY SIMILARITY.
CC ACT_SITE 217 217 BY SIMILARITY.
CC CONFLICT 183 183 E -> D (IN REF. 2).
CC CONFLICT 193 193 F -> S (IN REF. 2).
CC CONFLICT 201 201 T -> P (IN REF. 2).
CC CONFLICT 206 206 T -> P (IN REF. 2).
CC CONFLICT 230 230 K -> N (IN REF. 2).
CC SEQUENCE 272 AA; 30084 MW; 60DB1A8527815C05 CRC64;
Query Match 69.6%; Score 32; DB 1; Length 272;
Best Local Similarity 57.1%; Pred. No. 55;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 QKCHQK 7
Db 32 QKMCHRR 38
|||
|||

RESULT 8
TF2B_YEAST          STANDARD;          PRT;    345 AA.
ID TF2B_YEAST
AC P29055;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Transcription initiation factor IIB (TFIIB) (Transcription factor E).
GN SUA7 OR YPR086W OR P9513.4.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=92191276; PubMed=1547497;
RA Pinto I., Ware D.E., Hampsey M.;
RT "The yeast SUA7 gene encodes a homolog of human transcription factor
RT TFIIB and is required for normal start site selection in vivo.";
RL Cell 68:977-988(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC
CC -1- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
CC OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II.
CC -1- SUBUNIT: ASSOCIATES WITH TFIID-IIA (DA COMPLEX) TO FORM TFIID-
CC IIA-IIIB (DAB-COMPLEX) WHICH IS THEN RECOGNIZED BY POLYMERASE II.
CC -1- SUBCELLULAR LOCATION: Nuclear
CC -1- SIMILARITY: BELONGS TO THE TFIIB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M81380; AAA35126.1; -
CC DR EMBL: U51033; AAB68135.1; -
CC DR EMBL: S26707; S26707.
CC DR HSP: Q00403; IDL6.
CC DR TRANSFAC: T00819; -
CC DR SGD: S0006290; SUA7.
CC DR InterPro: IPR000553; Cyclin.
CC DR InterPro: IPR000812; TFIIB.
CC DR Pfam: PF00382; transcript_fac2; 2.
CC DR PRINTS: PR00685; TIFACTORIIB.
CC DR SMART: SM00385; CYCLIN; 2.
CC DR PROSITE: PS00782; TFIIB; 1.
CC DR Transcription regulation; Nuclear protein; Repeat; Zinc-finger.
CC KW
CC ZN_FING 24 48
CC REPEAT 136 212
CC REPEAT 242 318
CC FT REPEAT
CC FT REPEAT
CC SQ SEQUENCE 345 AA; 38200 MW; 8F1F6D24602436E2 CRC64;
Query Match 69.6%; Score 32; DB 1; Length 345;
Best Local Similarity 71.4%; Pred. No. 67;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 KLCHQKK 8
DB 155 KLCHDEK 161
||||:|
RESULT 9
Y303_MYCGE STANDARD; PRT; 357 AA.
AC P47545;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein MG303.
GN MG303.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
-----
RX MEDLINE=92191276; PubMed=1547497;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Boff K.F., Hu P.-C., Lucier T.S.,
RA Paterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U39710; AAC71525.1; -
CC DR TIGR: MG303; -
CC DR InterPro: IPR003593; AAA.
CC DR InterPro: IPR003439; ABC_transporter.
CC DR InterPro: IPR001687; ATP_GTP_A.
CC DR Pfam: PF00005; ABC_tran; 1.
CC DR SMART: SM00382; AAA; 1.
CC DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
CC KW Hypothetical protein; ATP-binding; Transport; Complete proteome.
CC FT NP_BIND 107 114
CC FT ATP (POTENTIAL).
CC SQ SEQUENCE 357 AA; 40786 MW; AFB1012F886E090E CRC64;
Query Match 69.6%; Score 32; DB 1; Length 357;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KLCHQK 7
DB 183 KLCHKK 188
||||:|
RESULT 10
PRIM_CHLTR STANDARD; PRT; 595 AA.
AC O84799;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA primase (EC 2.7.7.-).
GN DNAG OR CT794.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -1- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC -1- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE001351; AAC68389.1; -
 DR HSP; Q9X4D0; ID0Q.
 DR InterPro; IPR002936; Toprim.
 DR InterPro; IPR002694; ZnF_CHCC.
 DR Pfam; PF01751; Toprim; 1.
 DR Pfam; PF01807; zf-CHC2; 1.
 DR ProDom; PD002988; ZnF_CHCC; 1.
 DR SMART; SM00493; TOPRIN; 1.
 DR SMART; SM00400; ZnF_CHCC; 1.
 KW Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;
 KW Zinc-finger; Zinc; Metal-binding; Complete proteome.
 FT ZN_FING 38 62 CHC2-TYPE (BY SIMILARITY).
 SQ SEQUENCE 595 AA; 68037 MW; 536858EBAFCDFB6 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 595;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCQ 6
 :|||||
 Db 565 RKLCQ 570

RESULT 11

ID FETA_GORGO STANDARD; PRT; 609 AA.
 AC P28050;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-fetoprotein precursor (Alpha-fetoglobulin) (Alpha-1-fetoprotein).
 DE AFP.
 OS Gorilla gorilla gorilla (Lowland gorilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Gorilla.
 OX NCBI_TaxID=9595;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91169517; PubMed=1706310;
 RA Ryan S.C., Zielinski R., Dugaiczky A.;
 RT "Structure of the gorilla alpha-fetoprotein gene and the divergence of primates";
 RL Genomics 9:60-72(1991).
 CC -1- FUNCTION: BINDS COPPER, NICKEL, AND FATTY ACIDS AS WELL AS, AND BILIRUBIN LESS WELL THAN, SERUM ALBUMIN.
 CC -1- SUBUNIT: DIMERIC AND TRIMERIC FORMS HAVE BEEN FOUND IN ADDITION TO THE MONOMERIC FORM (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- PTM: SULFATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M38272; AAA73520.1; -
 DR PIR; A37970; FPGO.
 DR HSP; P02768; 1BJ5.
 DR InterPro; IPR000264; Serum_albumin.
 DR Pfam; PF00273; transport_prot; 3.
 DR PRINTS; PR00802; SERUMALBUMIN.
 DR ProDom; PD002486; Serum_albumin; 1.

DR SMART; SM00103; ALBUMIN; 3.
 DR PROSITE; PS00212; ALBUMIN; 2.
 KW Glycoprotein; Sulfation; Albumin; Plasma; Embryo; Repeat;
 KW Metal-binding; Copper; Nickel; Signal
 FT SIGNAL 1 18 BY SIMILARITY.
 FT CHAIN 19 609 ALPHA-FETOPROTEIN.
 FT METAL 22 22 COPPER AND NICKEL (BY SIMILARITY).
 FT DISULFID 99 114 BY SIMILARITY.
 FT DISULFID 113 124 BY SIMILARITY.
 FT DISULFID 148 193 BY SIMILARITY.
 FT DISULFID 192 201 BY SIMILARITY.
 FT DISULFID 224 270 BY SIMILARITY.
 FT DISULFID 269 277 BY SIMILARITY.
 FT DISULFID 289 303 BY SIMILARITY.
 FT DISULFID 302 313 BY SIMILARITY.
 FT DISULFID 384 393 BY SIMILARITY.
 FT DISULFID 416 462 BY SIMILARITY.
 FT DISULFID 461 472 BY SIMILARITY.
 FT DISULFID 485 501 BY SIMILARITY.
 FT DISULFID 500 511 BY SIMILARITY.
 FT DISULFID 538 583 BY SIMILARITY.
 FT DISULFID 582 591 BY SIMILARITY.
 FT CARBOHYD 251 251 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 609 AA; 68697 MW; E8AE548377DB60EB CRC64;

Query Match 69.6%; Score 32; DB 1; Length 609;
 Best Local Similarity 50.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCQKK 8
 :|||||:
 Db 96 EELCHEKE 103

RESULT 12

ID FETA_HUMAN STANDARD; PRT; 609 AA.
 AC P02771;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-fetoprotein precursor (Alpha-fetoglobulin) (Alpha-1-fetoprotein).
 DE AFP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83273654; PubMed=6192439;
 RA Morinaga T., Sakai M., Wegmann T.G., Tamaoki T.;
 RT "Primary structures of human alpha-fetoprotein and its mRNA";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4604-4608(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87185438; PubMed=2436661;
 RA Gibbs P.E.M., Zielinski R., Boyd C., Dugaiczky A.;
 RT "Structure, polymorphism, and novel repeated DNA elements revealed by a complete sequence of the human alpha-fetoprotein gene";
 RL Biochemistry 26:1332-1343(1987).
 RN [3]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=93278385; PubMed=7684942;
 RA Mcvey J.H., Michaelides K., Hansen L.P., Ferguson-Smith M., Tilghman S., Krumlauf R., Tuddenham E.G.D.;
 RT "A G->A substitution in an HNF I binding site in the human alpha-fetoprotein gene is associated with hereditary persistence of alpha-fetoprotein (HPAFP).";
 RL Hum. Mol. Genet. 2:379-379(1993).
 RN [4]
 RP SEQUENCE OF 429-556 FROM N.A.
 RX MEDLINE=83158778; PubMed=6187626;

RA Beattie W.G., Dugaiczkyk A.;
RT "Structure and evolution of human alpha-fetoprotein deduced from
RT partial sequence of cloned cDNA.";
RL Gene 20:415-422(1982).
RN [5]
RP PARTIAL SEQUENCE OF 19-609.
RX MEDLINE=91242409; PubMed=1709810;
RA Pucci P., Siciliano R., Malorni A., Marino G., Tecce M.F.,
RA Ceccarini C., Terrana B.;
RT "Human alpha-fetoprotein primary structure: a mass spectrometric
RT study.";
RL Biochemistry 30:5061-5066(1991).
RN [6]
RP PRELIMINARY SEQUENCE OF 19-35.
RX MEDLINE=77242506; PubMed=70228;
RA Yachnin S., Hsu R., Heinrikson R.L., Miller J.B.;
RT "Studies on human alpha-fetoprotein. Isolation and characterization
RT of monomeric and polymeric forms and amino-terminal sequence
RT analysis.";
RL Biochim. Biophys. Acta 493:418-428(1977).
RN [7]
RP PRELIMINARY SEQUENCE OF 19-38.
RX MEDLINE=78001760; PubMed=71198;
RA Aoyagi Y., Ikenaka T., Ichida F.;
RT "Comparative chemical structures of human alpha-fetoproteins from
RT fetal serum and from ascites fluid of a patient with hepatoma.";
RL Cancer Res. 37:3663-3667(1977).
RN [8]
RP PRELIMINARY SEQUENCE OF 19-39.
RX MEDLINE=75018719; PubMed=4138095;
RA Ruoslahti E., Pihko H., Vaheri A., Seppala M., Virolainen M.,
RA Kontinen A.;
RT "Alpha fetoprotein: structure and expression in man and inbred mouse
RT strains under normal conditions and liver injury.";
RL Johns Hopkins Med. J. Suppl. 3:249-255(1974).
RN [9]
RP GENE STRUCTURE.
RX MEDLINE=85182629; PubMed=2580830;
RA Sakai M., Morinaga T., Urano Y., Watanabe K., Wegmann T.G.,
RA Tanaaki T.;
RT "The human alpha-fetoprotein gene. Sequence organization and the 5'
RT flanking region.";
RL J. Biol. Chem. 260:5055-5060(1985).
RN [10]
RP METAL-BINDING.
RX MEDLINE=79001617; PubMed=80265;
RA Aoyagi Y., Ikenaka T., Ichida F.;
RT "Copper(II)-binding ability of human alpha-fetoprotein.";
RL Cancer Res. 38:3483-3486(1978).
RN [11]
RP BILIRUBIN-BINDING.
RX MEDLINE=80001710; PubMed=89900;
RA Aoyagi Y., Ikenaka T., Ichida F.;
RT "Alpha-Fetoprotein as a carrier protein in plasma and its bilirubin-
RT binding ability.";
RL Cancer Res. 39:3571-3574(1979).
RN [12]
RP SULFATION.
RX MEDLINE=86042625; PubMed=2414772;
RA Liu M.C., Yu S., Sy J., Redman C.M., Lipmann F.;
RT "Tyrosine sulfation of proteins from the human hepatoma cell line
RT HepG2.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7160-7164(1985).
CC -1- FUNCTION: BINDS COPPER, NICKEL, AND FATTY ACIDS AS WELL AS, AND
CC BILIRUBIN LESS WELL THAN, SERUM ALBUMIN. ONLY A SMALL PERCENTAGE
CC (LESS THAN 2%) OF THE HUMAN APP SHOWS ESTROGEN-BINDING PROPERTIES.
CC -1- SUBUNIT: DIMERIC AND TRIMERIC FORMS HAVE BEEN FOUND IN ADDITION
CC TO THE MONOMERIC FORM.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: PLASMA. SYNTHESIZED BY THE FETAL LIVER AND
CC YOLK SAC.
CC -1- DEVELOPMENTAL STAGE: OCCURS IN THE PLASMA OF FETUSES MORE THAN 4
CC WEEKS OLD, REACHES THE HIGHEST LEVELS DURING THE 12TH-16TH WEEK OF

CC GESTATION, AND DROPS TO TRACE AMOUNTS AFTER BIRTH. THE SERUM LEVEL
CC IN ADULTS IS USUALLY LESS THAN 40 NG/ML. APP OCCURS ALSO AT HIGH
CC LEVELS IN THE PLASMA AND ASCITIC FLUID OF ADULTS WITH HEPATOMA.
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- PTM: INDEPENDENT STUDIES SUGGEST HETEROGENEITY OF THE AMINO-
CC TERMINAL SEQUENCE OF THE MATURE PROTEIN AND OF THE CLEAVAGE SITE
CC OF THE SIGNAL SEQUENCE.
CC -1- PTM: SULFATED.
CC -1- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M10949; AAA51674.1; -;
DR EMBL; M10950; AAA51675.1; -;
DR EMBL; V01514; CAA24758.1; -;
DR EMBL; M16110; AAB58754.1; -;
DR EMBL; J00077; AAC95396.1; -;
DR EMBL; Z19532; CAA79592.1; -;
DR PIR; A03234; FPHU.
DR PIR; A26624; A26624.
DR HSSP; P02768; 1BJ5
DR GlycoSuiteDB; P02771; -;
DR Sienna-2DPAGE; P02771; -;
DR MIM; 104150; -;
DR InterPro; IPR000264; Serum_albumin.
DR Pfam; PF00273; transport_prot; 3.
DR PRINTS; PR00802; SERUMALBUMIN.
DR ProDom; PD002486; Serum_albumin; 1.
DR SMART; SM00103; ALBUMIN; 3.
DR PROSITE; PS00212; ALBUMIN; 2.
KW Glycoprotein; Sulfation; Albumin; Plasma; Embryo; Repeat;
KW Metal-binding; Copper; Nickel; Signal; Polymorphism.
FT SIGNAL 1 18
FT CHAIN 19 609 ALPHA-FETOPROTEIN.
FT METAL 22 22 COPPER AND NICKEL.
FT DISULFID 99 114
FT DISULFID 113 124
FT DISULFID 148 193
FT DISULFID 192 201
FT DISULFID 224 270
FT DISULFID 269 277
FT DISULFID 289 303
FT DISULFID 302 313
FT DISULFID 384 393
FT DISULFID 416 462
FT DISULFID 461 472
FT DISULFID 485 501
FT DISULFID 500 511
FT DISULFID 538 583
FT DISULFID 582 591
FT CARBOHYD 251 251
FT VARIANT 570 570 N-LINKED (GLCNAC. . .).
FT SEQUENCE 609 AA; 68677 MW; 4D4E45820E1C2D4F CRC64;
SQ
Query Match . 69.6%; Score 32; DB 1; Length 609;
Best Local Similarity 50.0%; Pred. NO. 1.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Oy 1 OKLCHKQK 8
Db 96 BELCHEKE 103
RESULT 13
Y081_CAEEL

```
ID Y081_CAEEEL STANDARD; PRT; 645 AA.
AC P34617;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical GTP-binding protein ZK1236.1 in chromosome III.
GN ZK1236.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durlin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC LEPA SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L13200; AAA28191.2; -
CC HSSP; P13551; 1ELO.
CC WormPep; ZK1236.1; CE01446.
CC InterPro; IPR000795; GTP_EFTU.
CC Pfam; PF004161; GTP_EFTU_D2.
CC Pfam; PF03144; GTP_EFTU_D2.
CC PROSITE; PS00301; EFATOR_GTP; 1.
KW Hypothetical protein; GTP-binding.
FT NP_BIND 49 56 GTP (POTENTIAL).
FT NP_BIND 108 112 GTP (POTENTIAL).
FT NP_BIND 162 165 GTP (POTENTIAL).
SQ SEQUENCE 645 AA; 72268 MW; 3F08EA3E5FD53819 CRC64;
Query Match 59.6%; Score 32; DB 1; Length 645;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 OKLCHQKK 8
Db 614 KKLHQKK 621
RESULT 14
SYL_MYCGE STANDARD; PRT; 895 AA.
ID SYL_MYCGE
AC P47587;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Isoleucyl-tRNA synthetase (EC 6.1.1.5) (Isoleucine--tRNA ligase)
DE (IleRS).
GN ILES OR MG345.
-----
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kellavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
[2]
RN SEQUENCE OF 262-371 AND 605-711 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -1- CATALYTIC ACTIVITY: ATP + L-isoleucine + tRNA(Ile) = AMP +
CC diphosphate + L-isoleucyl-tRNA(Ile).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U39716; AAC71570.1; -
CC EMBL; U02196; AAD12482.1; -
CC EMBL; U02254; AAD12519.1; -
CC HSSP; P41972; 1FFY.
CC TIGR; MG345; -
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR002301; tRNA-synt_1le.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00984; TRNASYNTHILE.
CC PROSITE; PS00178; AA-TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Zinc; Complete proteome.
FT SITE 57 67 "HIGH" REGION.
FT SITE 590 594 "KMSKS" REGION.
FT BINDING 593 593 ATP (BY SIMILARITY).
SQ SEQUENCE 895 AA; 104395 MW; 8C78DE6A05311B22 CRC64;
Query Match 69.6%; Score 32; DB 1; Length 895;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 OKLCHQ 6
Db 122 RKLCHQ 127
:|||||
RESULT 15
AT12_HUMAN STANDARD; PRT; 1593 AA.
ID AT12_HUMAN
AC P58397;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
```

DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=fetal lung;
RA MEDLINE=21264577; PubMed=11279086;
RA Cal S.; Argueelles J.M.; Fernandez P.L.; Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ250725; CAC20419.1; .
DR MIM: 606184; .
DR PROSITE: PS0215; ADAM_MEPPO; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE: PS00092; TSPI; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 545 596 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 SPACER 1.
FT DOMAIN 827 881 TSP TYPE-1 2.
FT DOMAIN 886 943 TSP TYPE-1 3.
FT DOMAIN 947 995 TSP TYPE-1 4.
FT DOMAIN 996 1315 SPACER 2.
FT DOMAIN 1316 1364 TSP TYPE-1 5.
FT DOMAIN 1367 1423 TSP TYPE-1 6.
FT DOMAIN 1426 1471 TSP TYPE-1 7.
FT DOMAIN 1426 1471 TSP TYPE-1 8.
FT DOMAIN 302 305 POLY-GLU.
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 393 393 BY SIMILARITY.
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).

FT	CARBOHYD	105	105	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	215	215	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	685	685	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	951	951	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1104	1104	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1275	1275	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1300	1300	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1371	1371	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1503	1503	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1593	AA; 177545	AA; 177545 MW; 079F48E63BD83A3	CRC64;

Query Match 69.6%; Score 32; DB 1; Length 1593;
Best Local Similarity 71.4%; Pred. No. 2.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHQK 7
Db 873 QKKCHEK 879

Search completed: November 1, 2002, 14:47:48
Job time : 2.41158 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:58 ; Search time 0.848875 Seconds
(without alignments)
905.569 Million cell updates/sec

Title: US-09-606-129A-19
Perfect score: 46
Sequence: 1 QKLCQKK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	100.0	295	2 A42268	billiverdin reducta
2	36	78.3	654	1 BVBYPI	CBPI protein - yea
3	34	73.9	208	2 E89046	protein cl08.2 [1
4	34	73.9	547	2 T39478	zinc-finger protei
5	34	73.9	1188	2 A71621	protein with 5'-3'
6	34	73.9	1295	2 T30528	reverse transcript
7	34	73.9	1473	2 T31422	C-terminal domain-
8	33	71.7	263	2 E72675	hypothetical prote
9	33	71.7	487	2 AF3147	glutamate-1-semial
10	33	71.7	476	2 D98140	hypothetical prote
11	33	71.7	673	2 S36336	probable transcrip
12	33	71.7	773	2 T00554	hypothetical prote
13	33	71.7	1247	2 T18671	hypothetical prote
14	32	69.6	135	1 W6WLEP	E6 protein - Europ
15	32	69.6	210	2 T16125	hypothetical prote
16	32	69.6	221	2 D90195	hypothetical prote
17	32	69.6	290	2 I39522	3-dehydroquinat d
18	32	69.6	345	1 S26707	transcription init
19	32	69.6	357	2 E64233	membrane transport
20	32	69.6	380	2 T24081	hypothetical prote
21	32	69.6	422	2 E96753	hypothetical prote
22	32	69.6	462	2 A46170	tektin A1 - sea ur
23	32	69.6	465	2 T27032	hypothetical prote
24	32	69.6	474	2 T00699	hypothetical prote
25	32	69.6	548	2 T22137	hypothetical prote
26	32	69.6	581	2 S44896	2k1236.1 protein -
27	32	69.6	595	2 F71471	probable DNA prima
28	32	69.6	609	1 FPHU	alpha-fetoprotein
29	32	69.6	609	1 FPGO	alpha-fetoprotein

30 32 69.6 895 2 B64238 isoleucine--trna l
31 32 69.6 920 2 JC7313 aryl hydrocarbon r
32 32 69.6 1056 2 T00060 hypothetical prote
33 31 67.4 54 2 H69960 exodeoxyribonuclea
34 31 67.4 170 1 UNOG neurotensin precur
35 31 67.4 251 2 G86368 hypothetical prote
36 31 67.4 266 2 T33411 hypothetical prote
37 31 67.4 271 2 T16421 hypothetical prote
38 31 67.4 299 2 A75401 ribulose-phosphate
39 31 67.4 324 2 G97141 probable phosphohy
40 31 67.4 343 2 AB1243 B. subtilis comg o
41 31 67.4 343 2 AF1605 B. subtilis comg o
42 31 67.4 364 2 T03892 hypothetical prote
43 31 67.4 436 2 D84782 probable proliline t
44 31 67.4 478 2 C71523 probable shikimate
45 31 67.4 491 2 G83850 ATP-dependent DNA

ALIGNMENTS

RESULT 1

A42268
billiverdin reductase (EC 1.3.1.24) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A42268
R:Fakhrail, H.; Maines, M.D.
J. Biol. Chem. 267, 4023-4029, 1992
A:Title: Expression and characterization of a cDNA for rat kidney billiverdin reductas
A:Reference number: A42268; MUID:92156147
A:Accession: A42268
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid; protein
A:Residues: 1-295 <FAK>
A:Cross-references: GB:M81681; NID:g203177; PIDN:AAA40830.1; PID:g203178
A:Experimental source: kidney
A:Note: sequence extracted from NCBI backbone (NCBIP:82800)
C:Keywords: liver; oxidoreductase

Query Match 100.0%; Score 46; DB 2; Length 295;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCQKK 8
DB 288 QKLCQKK 295

RESULT 2

BVBYPI
CBPI protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein HRA654; protein J0242; protein YJL209w
C:Species: Saccharomyces cerevisiae
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 12-Nov-1999
C:Accession: S05829; S50776; S56999; S45164
R:Dieckmann, C.L.; Homison, G.; Tzagoloff, A.
J. Biol. Chem. 259, 4732-4738, 1984
A:Title: Assembly of the mitochondrial membrane system. Nucleotide sequence of a yeast
A:Reference number: S05829; MUID:84185566
A:Accession: S05829
A:Molecule type: DNA
A:Residues: 1-654 <DE>
A:Cross-references: EMBL:K02647; NID:g171166; PIDN:AAA34474.1; PID:g171167
R:Vandenbol, M.; Durand, P.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.
Yeast 10, 1657-1662, 1994
A:Title: Sequence analysis of a 40.2 kb DNA fragment located near the left telomere o
A:Reference number: S50701; MUID:95242842
A:Accession: S50776
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-654 <VAV>
A:Cross-references: EMBL:Z34098; NID:g496934; PIDN:CAA84002.1; PID:g496953

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56835
A:Accession: S56999
A:Molecule type: DNA
A:Residues: 1-654 <VAN>
A:Cross-references: EMBL:Z49484; MID:g1015590; PIDN:CAA89506.1; PID:g1015591; GSPDB:GN000000
R:Purnelle, B.; Coster, F.; Goffeau, A.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56977
A:Accession: S56996
A:Molecule type: DNA
A:Residues: 637-654 <PUR>
A:Cross-references: EMBL:Z49484; GSPDB:GN00010; MIPS:YJL209W
C:Genetics:
A:Gene: SGD:CBP1; MIPS:YJL209W
A:Cross-references: SGD:S0003745; MIPS:YJL209W
A:Map position: 10L
A:Genome: nuclear
C:Function:
A:Description: pre-mRNA processing
A:Note: required for correct 5' terminal processing of cytochrome b pre-mRNA
C:Superfamily: CBP1 protein
C:Keywords: mitochondrion

Query Match 78.3%; Score 36; DB 1; Length 654;
Best Local Similarity 75.0%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKLCCHQK 8
:|||||
Db 514 KKLCHYKK 521

RESULT 3
E89046
protein C10G8.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E89046
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
A:Reference number: A75000; MID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999.
A:Accession: E89046
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-208 <STO>
A:Cross-references: GB:chr_V; PIDN:AAB09170.1; PID:g1572828; GSPDB:GN00023; CESP:C10G8.2
C:Genetics:
A:Gene: C10G8.2
A:Map position: 5

Query Match 73.9%; Score 34; DB 2; Length 208;
Best Local Similarity 83.3%; Pred. No. 54;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LCHQKK 8
:|||||
Db 172 MCHQKK 177

RESULT 4
T39478
zinc-finger protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: T39478
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
submitted to the EMBL Data Library, August 1997

A:Reference number: Z21858
A:Accession: T39478
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-547 <LYN>
A:Cross-references: EMBL:AL031349; PIDN:CAA20477.1; GSPDB:GN00067; SPDB:SPBC15D4.02
C:Experimental source: strain 972h-; cosmid c15D4
C:Genetics:
A:Gene: SPDB:SPBC15D4.02
A:Map position: 2
C:Superfamily: GAL4 zinc binuclear cluster homology
F:169-206/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 73.9%; Score 34; DB 2; Length 547;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCCHQK 7
:|||||
Db 427 RKLCHQK 433

RESULT 5
A71621
protein with 5'-3' exonuclease domain (Kem-1 family) PFB0205c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: A71621
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H. Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MID:99021743
A:Accession: A71621
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1188 <GAR>
A:Cross-references: GB:AE001380; GB:AE001362; MID:g3845120; PIDN:AAC71830.1; PID:g384
C:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0205c

Query Match 73.9%; Score 34; DB 2; Length 1188;
Best Local Similarity 71.4%; Pred. No. 2.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCCHQK 7
:|||||
Db 1117 EELCHQK 1123

RESULT 6
T30528
reverse transcriptase - rice blast fungus
C:Species: Magnaporthe grisea (rice blast fungus)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30528
R:Meyn III, M.A.; Farrall, L.; Valent, B.; Chumley, F.G.; Orbach, M.J.
submitted to the EMBL Data Library, August 1997
A:Description: Magnaporthe grisea repeated DNA element MGR583 is a member of the LINE
A:Reference number: Z20845
A:Accession: T30528
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1295 <MEY>
A:Cross-references: EMBL:AF018033; MID:g2454620; PID:g2454622; PIDN:AAB71689.1

Query Match 73.9%; Score 34; DB 2; Length 1295;
Best Local Similarity 71.4%; Pred. No. 2.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLCHQKK 8
:|||||

Db 33 ELCHQK 39

RESULT 7

T31422

C:terminal domain-binding protein rA9 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
R:Yuryev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Cord
Proc. Natl. Acad. Sci. U.S.A. 93, 6973-6980, 1996
A:Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts with
A:Reference number: Z21024; MUID:96293459
A:Accession: T31422
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1473 <YOR>
A:Cross-references: EMBL:U49057; NID:g1438533; PID:g1438534; PIDN:AAC52658.1
A:Experimental source: hippocampus

Query Match 73.9%; Score 34; DB 2; Length 1473;

Best Local Similarity 71.4%; Pred. No. 2.6e+02; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 1;

QY 1 QKLCHQK 7

|||||

Db 1425 QKICHSK 1431

RESULT 8

E72675

hypothetical protein APE0826 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: E72675
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339
A:Accession: E72675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-263 <KAW>
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79805.1; PID:g5104490
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0826
C:Superfamily: Aeropyrum pernix hypothetical protein APE0826

Query Match 71.7%; Score 33; DB 2; Length 263;

Best Local Similarity 83.3%; Pred. No. 99;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LCHQK 8

|||||

Db 258 LCHQKR 263

RESULT 9

AF3147

glutamate-1-semialdehyde 2,1-aminotransferase hemL [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AF3147
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193

A:Accession: AF3147

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-467 <KUR>

A:Cross-references: GB:AE008689; PIDN:AAL45596.1; PID:g17743315; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: hemL

A:Map position: linear chromosome

C:Superfamily: ornithine--oxo-acid aminotransferase

Query Match 71.7%; Score 33; DB 2; Length 467;

Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCHQ 6

|||||

Db 220 QKLCHQ 225

RESULT 10

D98140

hypothetical protein AGR_L159 [imported] - Agrobacterium tumefaciens (strain C58, Ce
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: D98140
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldm
A.; Liu, F.; Wolam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: D98140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-476 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK88646.1; PID:g15158369; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L159

A:Map position: linear chromosome

C:Superfamily: ornithine--oxo-acid aminotransferase

Query Match 71.7%; Score 33; DB 2; Length 476;

Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCHQ 6

|||||

Db 229 QKLCHQ 234

RESULT 11

S36336

probable transcription factor PLZF - human

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999

C:Accession: S36336; S31989

R:Chen, Z.; Brand, N.J.; Chen, A.; Chen, S.J.; Tong, J.H.; Wang, Z.Y.; Waxman, S.; Ze

EMBO J. 12, 1161-1167, 1993

A:Title: Fusion between a novel Krueppel-like zinc finger gene and the retinoic acid

A:Reference number: S36336; MUID:93209216

A:Accession: S36336

A:Molecule type: mRNA

A:Residues: 1-673 <CHE>

A:Cross-references: EMBL:Z19002; NID:g38517; PIDN:CAA79489.1; PID:g38518

C:Genetics:

A:Gene: PLZF

C:Superfamily: POZ domain homology

C:Keywords: zinc finger

F:20-118/Domain: POZ domain homology <POZ>

Query Match 71.7%; Score 33; DB 2; Length 673;

Best Local Similarity 83.3%; Pred. No. 2.1e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KLCHQK 7
:||||:
Db 605 KLCHQK 610

RESULT 12

T00554
hypothetical protein At2g39440 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F12L6.10
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00554; D84817
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.
A:Reference number: Z14168
A:Accession: T00554
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-773 <ROU>
A:Cross-references: EMBL:AC004218; NID:g3355463; PID:g3355473
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: D84817
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-773 <STO>
A:Cross-references: GB:AE002093; NID:g3355473; PIDN:ACC27835.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g39440; F12L6.10
A:Map position: 2
A:Introns: 35/1; 75/1; 117/1; 159/1; 222/3; 283/3; 294/1; 506/3; 567/1

Query Match 71.7%; Score 33; DB 2; Length 773;
Best Local Similarity 75.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHQK 8
:||||:
Db 462 KKLCHQK 469

RESULT 13

T18671
hypothetical protein B0240.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T18671
R:White, S.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19004
A:Accession: T18671
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1247 <WIL>
A:Cross-references: EMBL:Z74026; PIDN:CAA98*16.1; GSPDB:GN00023; CESP:B0240.2
C:Genetics:
A:Gene: CESP:B0240.2
A:Map position: 5
A:Introns: 29/2; 73/2; 108/2; 129/3; 202/2; 265/3; 401/3; 454/1; 466/3; 553/3; 594/2; 64

Query Match 71.7%; Score 33; DB 2; Length 1247;
Best Local Similarity 71.4%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KLCHQK 8
:||||:
Db 265 KLCHQK 271

RESULT 14

W6WLEP
E6 protein - European elk papillomavirus
C:Species: European elk papillomavirus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-May-2000
C:Accession: A29499; F94457; F94506
R:Ahola, H.; Bergman, P.; Stroem, A.C.; Moreno-Lopez, J.; Pettersson, U.
Gene 50, 195-205, 1986
A:Title: Organization and expression of the transforming region from the European elk
A:Reference number: A91567; MUID:87219878
A:Accession: A29499
A:Molecule type: DNA
A:Residues: 1-135 <AHO>
A:Cross-references: GB:M15953; NID:g333025; PIDN:AAA66849.1; PID:g484015
R:Eriksson, A.
unpublished results 1987, cited by GenBank
A:Reference number: A94457
A:Accession: F94457
A:Molecule type: DNA
A:Residues: 1-135 <ERI>
A:Cross-references: GB:M15953; NID:g333025; PIDN:AAA66849.1; PID:g484015
R:Pettersson, U.
submitted to GenBank, August 1987
A:Reference number: A94506
A:Accession: F94506
A:Molecule type: DNA
A:Residues: 1-135 <PFT>
A:Cross-references: GB:M15953; NID:g333025; PIDN:AAA66849.1; PID:g484015
C:Superfamily: papillomavirus E6 protein
C:Keywords: DNA binding; early protein; zinc finger

Query Match 69.6%; Score 32; DB 1; Length 135;
Best Local Similarity 71.4%; Pred. No. 88;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KLCHQK 8
:||||:
Db 24 KRCHEK 30

RESULT 15

T16125
hypothetical protein F21H12.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
C:Accession: T16125
R:Favella, T.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid F21H12.
A:Reference number: Z18464
A:Accession: T16125
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-210 <FAV>
A:Cross-references: EMBL:U23176; NID:g726404; PID:g726407; PIDN:AAC46715.1; CESP:F21H
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F21H12.3
A:Introns: 37/3; 62/2
C:Superfamily: Caenorhabditis elegans hypothetical protein F21H12.3

Query Match 69.6%; Score 32; DB 2; Length 210;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHQK 7
:||||:
Db 23 KNLCHQK 29

Search completed: November 1, 2002, 14:51:00
Job time : 4.01554 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run On: November 1, 2002, 14:44:48 ; Search time 0.66881 Seconds
(without alignments)
292.168 Million cell updates/sec

Title: us-09-606-129a-19
Perfect score: 46
Sequence: 1 QKLCHQKK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	69.6	197	2	US-08-377-309-3
2	32	69.6	197	4	US-09-186-723-3
3	32	69.6	197	5	PCT-US96-00996-6
4	32	69.6	389	2	US-08-377-309-6
5	32	69.6	389	4	US-09-186-723-6
6	32	69.6	389	5	PCT-US96-00996-9
7	32	69.6	590	2	US-08-377-309-2
8	32	69.6	590	4	US-09-186-723-2
9	32	69.6	590	5	PCT-US96-00996-5
10	32	69.6	609	1	US-08-222-619-4
11	32	69.6	609	5	PCT-US95-04075-4
12	31	67.4	169	4	US-09-002-114-3
13	30	65.2	160	1	US-07-847-010-11
14	30	65.2	160	1	US-07-847-010-17
15	30	65.2	457	1	US-08-206-006-2
16	30	65.2	474	4	US-09-538-414-8
17	30	65.2	1104	2	US-08-327-832-5
18	30	65.2	1104	2	US-08-828-584-5
19	29	63.0	16	4	US-09-146-580-4
20	29	63.0	18	4	US-09-316-630-3
21	29	63.0	18	4	US-09-316-630-4
22	29	63.0	72	3	US-08-338-579A-101
23	29	63.0	72	3	US-08-338-579A-102
24	29	63.0	83	4	US-09-442-631-2
25	29	63.0	104	4	US-09-146-580-2
26	29	63.0	127	4	US-09-146-580-1
27	29	63.0	149	3	US-08-338-579A-95

28	29	63.0	308	1	US-07-828-700-9	Sequence 9, Appl1
29	29	63.0	562	2	US-08-973-675-2	Sequence 2, Appl1
30	29	63.0	664	4	US-09-268-140-2	Sequence 2, Appl1
31	29	63.0	711	3	US-08-772-270A-12	Sequence 12, Appl1
32	29	63.0	729	2	US-08-677-298-2	Sequence 2, Appl1
33	29	63.0	781	1	US-08-280-690-2	Sequence 2, Appl1
34	28	60.9	57	4	US-09-103-478-23	Sequence 23, Appl1
35	28	60.9	57	4	US-09-193-931C-23	Sequence 23, Appl1
36	28	60.9	77	4	US-09-262-773-207	Sequence 207, App
37	28	60.9	90	4	US-09-103-478-21	Sequence 21, Appl1
38	28	60.9	90	4	US-09-103-478-22	Sequence 22, Appl1
39	28	60.9	90	4	US-09-103-478-24	Sequence 24, Appl1
40	28	60.9	90	4	US-09-103-478-25	Sequence 25, Appl1
41	28	60.9	90	4	US-09-193-931C-21	Sequence 21, Appl1
42	28	60.9	90	4	US-09-193-931C-22	Sequence 22, Appl1
43	28	60.9	90	4	US-09-193-931C-24	Sequence 24, Appl1
44	28	60.9	90	4	US-09-193-931C-25	Sequence 25, Appl1
45	28	60.9	133	4	US-09-262-773-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-08-377-309-3
; Sequence 3, Application US/08377309A
; Patent No. 5965528
; GENERAL INFORMATION:
; APPLICANT: Murgita, Robert A.
; TITLE OF INVENTION: RECOMBINANT HUMAN ALPHA-FETOPROTEIN AS
; FILE REFERENCE: 06727/005001
; CURRENT APPLICATION NUMBER: US/08/377,309A
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-377-309-3

Query Match 69.6%; Score 32; DB 2; Length 197;
Best Local Similarity 50.0%; Pred. No. 1.le+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKLCHQKK 8
Db 77 EELCHEKE 84
:::|::|:

RESULT 2
US-09-186-723-3
; Sequence 3, Application US/09186723
; Patent No. 6288034
; GENERAL INFORMATION:
; APPLICANT: Murgita, Robert A.
; TITLE OF INVENTION: RECOMBINANT HUMAN ALPHA-FETOPROTEIN AS
; FILE REFERENCE: 06727/005002
; CURRENT APPLICATION NUMBER: US/09/186,723
; CURRENT FILING DATE: 1998-11-05
; EARLIER APPLICATION NUMBER: 08/377,309
; EARLIER FILING DATE: 1995-01-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-186-723-3

Query Match 69.6%; Score 32; DB 4; Length 197;


```
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00996
; FILING DATE: 24-JAN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/377,317
; FILING DATE: 24-JAN-1995
; CLASSIFICATION:
; APPLICATION NUMBER: 08/377,311
; FILING DATE: 24-JAN-1995
; CLASSIFICATION:
; APPLICATION NUMBER: 08/377,309
; FILING DATE: 24-JAN-1995
; CLASSIFICATION:
; APPLICATION NUMBER: 08/377,316
; FILING DATE: 24-JAN-1995
; CLASSIFICATION:
; APPLICATION NUMBER: 08/505,012
; FILING DATE: 21-JULY-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06727/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 389 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-00996-9

Query Match 69.6%; Score 32; DB 5; Length 389;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 OKLCHQKK 8
Db 77 EELCHEKE 84

RESULT 7
US-08-377-309-2
; Sequence 2, Application US/08377309A
; Patent No. 5965528
; GENERAL INFORMATION:
; APPLICANT: Murgita, Robert A.
; TITLE OF INVENTION: RECOMBINANT HUMAN ALPHA-FETOPROTEIN AS
; FILE REFERENCE: 06727/005001
; CURRENT APPLICATION NUMBER: US/08/377,309A
; CURRENT FILING DATE: 1995-01-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-377-309-2

Query Match 69.6%; Score 32; DB 2; Length 590;
Best Local Similarity 50.0%; Pred. No. 3e+02;
```

```
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 OKLCHQKK 8
Db 77 EELCHEKE 84

RESULT 8
US-09-186-723-2
; Sequence 2, Application US/09186723
; Patent No. 6288034
; GENERAL INFORMATION:
; APPLICANT: Murgita, Robert A.
; TITLE OF INVENTION: RECOMBINANT HUMAN ALPHA-FETOPROTEIN AS
; FILE REFERENCE: 06727/005002
; CURRENT APPLICATION NUMBER: US/09/186,723
; CURRENT FILING DATE: 1998-11-05
; EARLIER APPLICATION NUMBER: 08/377,309
; EARLIER FILING DATE: 1995-01-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-186-723-2

Query Match 69.6%; Score 32; DB 4; Length 590;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 OKLCHQKK 8
Db 77 EELCHEKE 84

RESULT 9
PCT-US96-00996-5
; Sequence 5, Application PC/TUS9600996
; GENERAL INFORMATION:
; APPLICANT: Murgita, Robert A.
; TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF CLONED
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street, Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00996
; FILING DATE: 24-JAN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/377,317
; FILING DATE: 24-JAN-1995
; CLASSIFICATION:
; APPLICATION NUMBER: 08/377,311
; FILING DATE: 24-JAN-1995
; CLASSIFICATION:
; APPLICATION NUMBER: 08/377,309
; FILING DATE: 24-JAN-1995
; CLASSIFICATION:
; APPLICATION NUMBER: 08/377,316
; FILING DATE: 24-JAN-1995
```

CLASSIFICATION:
APPLICATION NUMBER: 08/505,012
FILING DATE: 21-JULY-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06727/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 590 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-00996-5

Query Match 59.6%; Score 32; DB 5; Length 590;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCQKK 8
:|||||:
Db 77 EELCHEKE 84

RESULT 10
US-08-222-619-4
Sequence 4, Application US/08222619
Patent No. 5652352
GENERAL INFORMATION:
APPLICANT: Lichenstein, Henri
APPLICANT: Lyons, David
APPLICANT: Wurfel, Mark
APPLICANT: Wright, Samuel
TITLE OF INVENTION: Afamin: A Human Serum Albumin-Like
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Center, Patent Operations/RRC
STREET: 1840 DeHavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: U.S.
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,619
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 609 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-222-619-4

Query Match 69.6%; Score 32; DB 1; Length 609;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCQKK 8
:|||||:

Db 96 EELCHEKE 103
RESULT 11
PCT-US95-04075-4
Sequence 4, Application PC/TUS9504075
GENERAL INFORMATION:
APPLICANT: AMGEN INC.
TITLE OF INVENTION: Afamin: A Human Serum Albumin-Like
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Center, Patent Operations/RRC
STREET: 1840 DeHavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: U.S.
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04075
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 609 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US95-04075-4

Query Match 69.6%; Score 32; DB 5; Length 609;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCQKK 8
:|||||:
Db 96 EELCHEKE 103

RESULT 12
US-09-002-114-3
Sequence 3, Application US/09002114
Patent No. 6274720
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN PREPRONEUROTENSIN/NEUROMEDIN N
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/002,114
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0450 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 169 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 163424

US-09-002-114-3

Query Match 67.4%; Score 31; DB 4; Length 169;

Best Local Similarity 57.1%; Pred. No. 1.5e+02;

Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHK 7

|||:

Db 107 QKICHSR 113

RESULT 13

US-07-847-010-11

; Sequence 11, Application US/07847010

; Patent No. 5693495

; GENERAL INFORMATION:

; APPLICANT: Breiteneder, Heimo

; APPLICANT: Reikerstorfer, Arnold

; APPLICANT: Valenta, Rudolf

; APPLICANT: Hoffmann - Sommergruber, Karin

; APPLICANT: Breitenbach, Michael

; APPLICANT: Kraft, Dietrich

; APPLICANT: Rumpold, Heimit

; APPLICANT: Scheiner, Otto

; APPLICANT: Ebner, Christof

; APPLICANT: Ferreira, Fatima

; TITLE OF INVENTION: Allergens of Alder Pollen and

; APPLICATION INFORMATION:

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/847,010

; FILING DATE: 01-JUN-1992

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jones III, Harry C

; REGISTRATION NUMBER: 20,280

; REFERENCE/DOCKET NUMBER: 6530-010

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 160 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: hazel (Corylus sp.)

; IMMEDIATE SOURCE:

; LIBRARY: POLLEN FROM ALLERGEN AB, ENGELHOLM, SWEDEN

US-07-847-010-11

Query Match 65.2%; Score 30; DB 1; Length 160;

Best Local Similarity 50.0%; Pred. No. 2.1e+02;

Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKLCHK 8

|||:

Db 97 EKVCHEK 104

RESULT 14

US-07-847-010-17

; Sequence 17, Application US/07847010

; Patent No. 5693495

; GENERAL INFORMATION:

; APPLICANT: Breiteneder, Heimo

; APPLICANT: Reikerstorfer, Arnold

; APPLICANT: Valenta, Rudolf

; APPLICANT: Hoffmann - Sommergruber, Karin

; APPLICANT: Breitenbach, Michael

; APPLICANT: Kraft, Dietrich

; APPLICANT: Rumpold, Heimit

; APPLICANT: Scheiner, Otto

; APPLICANT: Ebner, Christof

; APPLICANT: Ferreira, Fatima

; TITLE OF INVENTION: Allergens of Alder Pollen and

; APPLICATION INFORMATION:

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/847,010

; FILING DATE: 01-JUN-1992

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jones III, Harry C

; REGISTRATION NUMBER: 20,280

; REFERENCE/DOCKET NUMBER: 6530-010

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 160 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: hazel (Corylus sp.)

; IMMEDIATE SOURCE:

; LIBRARY: POLLEN FROM ALLERGEN AB, ENGELHOLM, SWEDEN

US-07-847-010-17

Query Match 65.2%; Score 30; DB 1; Length 160;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKLCHQKK 8
:|:|:|:
Db 97 EKVCHLKL 104

RESULT 15

US-08-206-006-2
; Sequence 2, Application US/08206006
; Patent No. 5474912
; GENERAL INFORMATION:
; APPLICANT: Sherman, David H.
; APPLICANT: Hu, Wei-Shou
; APPLICANT: Malmberg, Li-Hong
; TITLE OF INVENTION: Method for Increasing Production of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lorraine R. Sherman
; STREET: 3201 Cavell Lane
; CITY: Minneapolis
; STATE: Minnesota
; COUNTRY: U.S.A.
; ZIP: 55426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,006
; FILING DATE: 03-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherman, Lorraine R.
; REGISTRATION NUMBER: 30,105
; REFERENCE/DOCKET NUMBER: 6102
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-206-006-2

Query Match 65.2%; Score 30; DB 1; Length 457;
Best Local Similarity 66.7%; Pred. No. 5.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKLCHQ 6
:|:|:|:
Db 262 QRLCHE 267

Search completed: November 1, 2002, 14:51:58
Job time: 1.66881 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:40:53 ; Search time 1.8135 Seconds
(without alignments)
489.985 Million cell updates/sec

Title: US-09-606-129A-19
Perfect score: 46
Sequence: 1 QKLC HQKK 8
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	82.6	141	22	AB12078 Human 23 kD basic
2	35	78.3	91	22	AA875580 Human secreted pro
3	35	76.1	301	22	ABG07163 Novel human diagno
4	35	76.1	551	20	AA129955 Mouse CG1CE protei
5	34	73.9	60	22	AA484701 Human immune/haema
6	34	73.9	88	21	AA856915 Human prostate can
7	34	73.9	102	22	AA876544 Human colon cancer
8	34	73.9	104	22	AA012476 Human polypeptide
9	34	73.9	522	22	AA95513 Human protein sequ
10	34	73.9	790	22	AA93045 Human protein sequ
11	34	73.9	1188	21	AA818183 Plasmodium falcipa

12	33	71.7	51	22	AA87820 Human immune/haema
13	33	71.7	87	22	AA83632 Human immune/haema
14	33	71.7	112	22	AA878167 Human phosphatidic
15	33	71.7	196	22	ABG22874 Novel human diagno
16	33	71.7	197	21	AA808675 Arabidopsis thailia
17	33	71.7	210	22	ABG23740 Novel human diagno
18	33	71.7	220	22	ABG11715 Novel human diagno
19	33	71.7	242	21	AA808674 Arabidopsis thailia
20	33	71.7	305	22	ABG20204 Novel human diagno
21	33	71.7	435	22	ABG27718 Novel human diagno
22	33	71.7	678	22	AA825801 Human protein sequ
23	33	71.7	778	21	AA879180 Haematopoietic ste
24	32	69.6	23	19	AA879501 Wild-type yeast tr
25	32	69.6	23	19	AA879502 Loss-of-function m
26	32	69.6	23	19	AA879503 Loss-of-function m
27	32	69.6	23	19	AA879504 Loss-of-function m
28	32	69.6	23	19	AA879505 Loss-of-function m
29	32	69.6	23	19	AA879506 Loss-of-function m
30	32	69.6	23	19	AA879507 Loss-of-function m
31	32	69.6	23	19	AA879508 Loss-of-function m
32	32	69.6	23	19	AA879509 Loss-of-function m
33	32	69.6	23	19	AA879510 Human brain expres
34	32	69.6	33	22	AA859485 Human bone marrow
35	32	69.6	33	22	AB842370 Peptide #9876 enco
36	32	69.6	55	22	AB825845 Protein #7844 enco
37	32	69.6	55	22	AA863258 Human brain expres
38	32	69.6	55	22	AA876070 Human bone marrow
39	32	69.6	55	22	AA836177 Peptide #10214 enc
40	32	69.6	55	22	AA816899 Human nervous syst
41	32	69.6	68	22	AA806457 Human polypeptide
42	32	69.6	87	22	AA828959 Peptide #1610 enco
43	32	69.6	88	22	AB834127 Peptide #1633 enco
44	32	69.6	88	22	AB834127 Peptide #1633 enco
45	32	69.6	88	22	AB819568 Protein #1567 enco

ALIGNMENTS

RESULT 1
AB12078
ID AB12078 standard; peptide: 141 AA.
XX
AC AB12078;
XX
DT 11-JAN-2002 (first entry)
XX
XX Human 23 kD basic protein homologue, SEQ ID NO:2448.
DE
DE Human: cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibitor; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytosstatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer.
XX
XX Homo sapiens.
OS
XX WO200157188-A2.
PN
XX 09-AUG-2001.
PD
XX 05-FEB-2001; 2001WO-US03800.
PF
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
PR

XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT;
 PI WPT; 2001-457740/49.
 XX N-PSDB; ABA09322.
 DR Human proteins and DNA encoding sequences useful for preventing,
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer -
 XX
 XX Claim 20; Page 305; 1963pp; English.
 XX
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention.
 XX
 SQ Sequence 141 AA;
 Query Match 82.6%; Score 38; DB 22; Length 141;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QKLCQKK 8
 |||:||||
 Db 134 QKTCYQKK 141
 RESULT 2
 AAB75580
 ID AAB75580 standard; Protein; 91 AA.
 XX AC
 XX AAB75580;
 XX
 XX 06-APR-2001 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 28 SEQ ID NO:134.
 XX

KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
 KW ophthalmological; vulnery; autoimmune disease; cardiovascular disorder;
 KW hyperproliferative disorders; cerebrovascular disorder; wound healing;
 KW nervous system disorder; ocular disorder; skin aging; chemotaxis;
 KW food additive.
 XX
 OS Homo sapiens.
 XX
 XX WO200077026-A1.
 PN
 XX 21-DEC-2000.
 PD
 XX 01-JUN-2000; 2000WO-US14973.
 PF
 XX 11-JUN-1999; 99US-0138630.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 PI Rosen CA, Ruben SM, Komatsoulis GA;
 PI
 XX WPI; 2001-071258/08.
 DR N-PSDB; AAF64203.
 DR
 XX Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers -
 XX
 PS Disclosure; Page 60; 542pp; English.
 XX
 CC Human secreted proteins AAB75506 - AAB75554 are encoded by polynucleotide
 CC sequences AAF64176 - AAF64224. The specification includes amino acid
 CC secreted proteins, and protein sequences with which they share homology.
 CC The proteins and polynucleotides, their agonists and antagonists have
 CC activities dependent on the tissues and cells in which they are
 CC expressed, examples of these activities include, immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
 CC virucide; fungicide; ophthalmological; and vulnerary. The proteins,
 CC polynucleotides, agonists and antagonists can be used to treat or detect
 CC or diagnose various diseases and disorders including, autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
 CC e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin ageing due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities. Included in the invention are
 CC polynucleotide sequences AAF64167 - AAF64175 and peptide AAB75505 which
 CC are used in the isolation, identification and characterisation of the
 CC proteins of the invention.
 XX
 SQ Sequence 91 AA;
 Query Match 78.3%; Score 36; DB 22; Length 91;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QKLCQK 6
 |||||
 Db 30 QKLCQK 35
 RESULT 3
 ABG07163
 ID ABG07163 standard; Protein; 301 AA.

XX AC ABG07163;
XX DT 13-FEB-2002 (first entry)
XX DE
XX DE Novel human diagnostic protein #7154.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS71350.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 37522; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 301 AA;
Query Match 76.1%; Score 35; DB 22; Length 301;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 QKLCHQKK 8
Db 134 EKLCHSRK 141
RESULT 4
AAV29955
ID AAV29955 standard; Protein; 551 AA.
XX AAV29955;
AC

XX DT 22-NOV-1999 (first entry)
XX DE Mouse CGICE protein sequence.
XX KW CGICE; Best's macular dystrophy; mutation; diagnosis; detection;
XX KW BMD; age-related macular dystrophy.
XX OS Mus sp.
XX PN WO9943695-A1.
XX PD 02-SEP-1999.
XX PF 22-FEB-1999; 99WO-US03790.
XX PR 25-FEB-1998; 98US-0075941.
XX PR 18-DEC-1998; 98US-0112926.
XX PA (MERI) MERCK & CO INC.
XX PA (UYUP-) UNIV UPPSALA.
XX PI Petrukhin K, Caskey CT, Metzker M, Wadelius C;
XX DR WPI; 1999-540560/45.
XX DR N-PSDB; AAZ21229.
XX PT Human and mouse polynucleotides encoding CGICE polypeptides -
XX PS Claim 7; Fig 8; 67pp; English.
XX CC The present sequence represents the mouse CGICE protein. When the CGICE
XX CC gene is mutated it is responsible for Best's macular dystrophy (BMD).
XX CC Polynucleotides encoding CGICE are useful for diagnosing whether a
XX CC patient carries a mutation in the CGICE gene. Normal and mutated
XX CC CGICE proteins are useful for identifying activators and/or inhibitors
XX CC of these proteins, in order to treat BMD. The CGICE gene offers a
XX CC simpler and cheaper method of diagnosing BMD without the need for the
XX CC presence of the patient. The gene may also be useful to discovering
XX CC the genetic cause of age-related macular dystrophy.
XX SQ Sequence 551 AA;
Query Match 76.1%; Score 35; DB 20; Length 551;
Best Local Similarity 62.5%; Pred. No. 2.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 QKLCHQKK 8
Db 490 QEICHMKK 497
RESULT 5
AAM84701
ID AAM84701 standard; Protein; 60 AA.
XX AAM84701;
XX AC AAM84701;
XX DT 07-NOV-2001 (first entry)
XX DE Human immune/haematopoietic antigen SEQ ID NO:12294.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01354.
XX

PT metastasis -

XX Claim 11; SEQ ID NO 12294; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.

XX Sequence 60 AA;

Query Match 73.9%; Score 34; DB 22; Length 60;
 Best Local Similarity 71.4%; Pred. No. 46;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 KLCHQKK 8

Db 47 RLCHKKK 53

RESULT 6
 AAB56915

ID AAB56915 standard; Protein: 88 AA.

XX AAB56915;

XX 13-MAR-2001 (first entry)

DE Human prostate cancer antigen protein sequence SEQ ID NO:1493.

XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease.

XX Homo sapiens.

XX WO200055174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587513/55.

XX N-PSDB; AAF16118.

XX Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -

XX Claim 11; Page 1930-1931; 2338pp; English.

XX

CC AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention.

XX Sequence 88 AA;

Query Match 73.9%; Score 34; DB 21; Length 88;
 Best Local Similarity 62.5%; Pred. No. 65;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 QKLCHQKK 8

Db 62 QELCHQOE 69

RESULT 7
 AAG76544

ID AAG76544 standard; Protein: 102 AA.

XX AC AAG76544;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:7308.

DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX N-PSDB; AAH35949.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 PT Claim 11; Page 8741-8742; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.

CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 102 AA;

Query Match 73.9%; Score 34; DB 22; Length 102;
Best Local Similarity 83.3%; Pred. No. 74;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LCHQKK 8
:|||||
Db 61 ICHQKK 66

RESULT 8
AAO12476
ID AAO12476 standard; Protein; 104 AA.

XX AC AAO12476;
XX AC
DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 26368.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX N-PSDB; AA192407.

XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX

PS Claim 20; SEQ ID NO 26368; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
X.

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 104 AA;

Query Match 73.9%; Score 34; DB 22; Length 104;
Best Local Similarity 71.4%; Pred. No. 76;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 KLCHQKK 8
:|||||
Db 60 RFCHQKK 66

RESULT 9
AAB95513
ID AAB95513 standard; Protein; 522 AA.

XX AC AAB95513;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:18081.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX

PS Claim 8; SEQ ID 18081; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

Sequence 522 AA;
Query Match 73.9%; Score 34; DB 22; Length 522;
Best Local Similarity 62.5%; Pred. No. 3.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 OKLCHOKK 8
I:||||:
Db 157 QELCHQOE 164

RESULT 10
AAB93045
ID AAB93045 standard; Protein: 790 AA.

XX AC AAB93045;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:11834.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EPI074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -

XX PS Claim 8; SEQ ID 11834; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesising polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

Sequence 790 AA;

Query Match 73.9%; Score 34; DB 22; Length 790;
Best Local Similarity 62.5%; Pred. No. 4.7e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 OKLCHOKK 8
I:||||:
Db 157 QELCHQOE 164

RESULT 11

AAB18183
ID AAB18183 standard; Protein: 1188 AA.

XX AC AAB18183;

XX DT 07-NOV-2000 (first entry)

XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:40.

XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
antimalarial; malaria; protozoacide; infection; insecticide.

XX OS Plasmodium falciparum.

XX PN WO200025728-A2.

XX PD 11-MAY-2000.

XX PF 05-NOV-1999; 99WO-US26796.

XX PR 05-NOV-1998; 98US-0107131.

XX PA (HOFF/) HOFFMAN S.

XX PA (CARU/) CARUCCI D.

XX PA (GARD/) GARDNER M.

XX PA (VENT/) VENTER J C.

XX PI Hoffman S, Carucci D, Gardner M, Venter JC;

XX DR WPI; 2000-365347/31.

XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
Plasmodium falciparum, useful as antimalarial vaccines and in the
diagnosis of P.falciparum infection -

XX PS Disclosure: Page 101-104; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded
by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAH70078 to AAH70287 and AAH18144 to AAH18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not

CC specifically mentioned within the specification.

XX Sequence 1188 AA;

Query Match 73.9%; Score 34; DB 21; Length 1188;

Best Local Similarity 71.4%; Pred. No. 6.8e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 OKLCHQK 7

:|||||

Db 1117 EELCHQK 1123

RESULT 12

AAAM87820

ID AAAM87820 standard; Protein; 51 AA.

XX AC AAAM87820;

XX DT 07-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen SEQ ID NO:15413.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
cytostatic; gene therapy; vaccine; metastasis.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01354.

XX PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0225759.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226686.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0227009.

PR 01-SEP-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPT; 2001-483426/52.
DR N-PSDB; AAK60601.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Claim 11; SEQ ID NO 15413; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 51 AA;
SQ
Query Match 71.7%; Score 33; DB 22; Length 51;
Best Local Similarity 62.5%; Pred. No. 59;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 QKLCHQKK 8
II:::II
Db 15 QKICNKKK 22
RESULT 13
AAM83632
ID AAM83632 standard; Protein; 87 AA.
XX
XX AAM83632;
AC
XX

DT 07-NOV-2001 (first entry)
XX Human immune/hematopoietic antigen SEQ ID NO:11225.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.

```
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR N-PSDB; AAK56413.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Claim 11: SEQ ID NO 11225; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 87 AA;
Query Match 71.7%; Score 33; DB 22; Length 87;
Best Local Similarity 71.4%; Pred. No. 96;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 KLCHQKK 8
Db 76 QLCHKKK 82
:|:|:|
RESULT 14
AAG78167
ID AAG78167 standard; Protein; 112 AA.
XX
AC AAG78167;
XX
XX 04-DEC-2001 (first entry)
XX
XX Human phosphatidic acid phosphatase 2-12.
XX
XX Human; phosphatidic acid phosphatase 2-12; cytostatic; virucidal;
XX immunomodulatory; antiinflammatory; haemostatic; malignant tumour;
XX human immunodeficiency virus; HIV; infection; immunological disease;
XX gene therapy.
XX
XX Homo sapiens.
XX
XX WO200173055-A1.
XX
XX 04-OCT-2001.
XX
XX 26-MAR-2001; 2001WO-CN00478.
XX
XX 28-MAR-2000; 2000CN-0115197.
XX
```

PA (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.

XX Mao Y, Xie Y;

XX WPI: 2001-597124/67.

DR N-PSDB: AAI64606.

XX Human phosphatidic acid phosphatase 2-12 and encoded polynucleotide,
XX used in diagnosis and treatment of malignant tumours, haemopathy, human
XX immunodeficiency virus infection, immunological diseases and
XX inflammation -

PS Claim 1; Page 31; 35pp: Chinese.

XX The invention relates to the human phosphatidic acid phosphatase 2-12
XX with cytostatic, virucidal, immunomodulatory, antiinflammatory and
XX haemostatic activity. The protein and encoding polynucleotide are used in
XX diagnosis and treatment of malignant tumour, haemopathy, human
XX immunodeficiency virus (HIV) infection, immunological diseases and
XX various inflammations. The polynucleotide is useful in gene therapy.

XX Sequence 112 AA;

Query Match 71.7%; Score 33; DB 22; Length 112;

Best Local Similarity 50.0%; Pred. No. 1.2e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 OKLCHQKK 8

DB :|:|:|

91 EKICHKEK 98

RESULT 15

ABG22874

ID ABG22874 standard; Protein; 196 AA.

XX AC

XX AC

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #22865.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB: AAS87061.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID NO 53233; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 196 AA;

Query Match 71.7%; Score 33; DB 22; Length 196;

Best Local Similarity 83.3%; Pred. No. 2e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 OKLCHQ 6

DB ||:|:|

2 QKVCHQ 7

Search completed: November 1, 2002, 14:47:08

Job time : 3.8135 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:08 ; Search time 1.2717 Seconds
(without alignments)
952.238 Million cell updates/sec

Title: US-09-606-129a-34
Perfect score: 40
Sequence: 1 KRRILHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_l9:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_cheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	40	100.0	235	11 Q9CY64	Q9cy64 mus musculus
2	40	100.0	296	4 Q9BRW8	Q9brw8 homo sapien
3	40	100.0	296	4 Q96QL4	Q96ql4 homo sapien
4	40	100.0	303	11 Q9DD21	Q9dd21 mus musculus
5	38	95.0	512	10 Q9LML0	Q9lml0 arabidopsis
6	38	95.0	599	10 Q9FHV3	Q9fhv3 arabidopsis
7	38	95.0	633	10 Q49511	Q49511 arabidopsis
8	36	90.0	279	11 Q9QXV3	Q9qxv3 mus musculus
9	36	90.0	508	11 Q9P6S2	Q9p6s2 mus musculus
10	35	87.5	267	2 Q9AKS0	Q9aks0 pseudomonas
11	35	87.5	495	10 Q9ST63	Q9st63 solanum tub
12	35	87.5	499	10 Q941V1	Q941v1 oryza sativ
13	35	87.5	508	10 Q90874	Q90874 arabidopsis
14	34	85.0	106	5 Q9XZM5	Q9xzm5 leishmania
15	34	85.0	111	4 Q9H007	Q9h007 homo sapien
16	34	85.0	233	4 Q9UIJ2	Q9uij2 homo sapien

17	34	85.0	235	4 Q9P0U6	Q9p0u6 homo sapien
18	34	85.0	279	4 Q43658	Q43658 homo sapien
19	34	85.0	279	4 Q9HD98	Q9hd98 homo sapien
20	34	85.0	279	4 Q9UK52	Q9uk52 homo sapien
21	34	85.0	283	4 Q9HD99	Q9hd99 homo sapien
22	34	85.0	294	4 Q00532	Q00532 homo sapien
23	34	85.0	338	3 Q9CIL3	Q9cil3 neurospora
24	34	85.0	422	4 Q9UK53	Q9uk53 homo sapien
25	34	85.0	422	4 Q9UIJ4	Q9uij4 homo sapien
26	34	85.0	455	5 Q9VJ28	Q9vj28 drosophila
27	34	85.0	497	10 Q9C8V1	Q9c8v1 arabidopsis
28	33	82.5	137	4 Q9NSR1	Q9nsr1 homo sapien
29	33	82.5	398	5 Q9VT97	Q9vt97 drosophila
30	33	82.5	401	10 Q9M149	Q9m149 arabidopsis
31	33	82.5	431	10 Q04613	Q04613 arabidopsis
32	33	82.5	601	4 Q9BRB5	Q9brb5 homo sapien
33	33	82.5	879	13 Q90ZT8	Q90zt8 gallus gall
34	32	80.0	102	10 Q9S738	Q9s738 lycopersico
35	32	80.0	112	11 Q9D4I7	Q9d4i7 mus musculus
36	32	80.0	189	2 Q03949	Q03949 anabena sp
37	32	80.0	193	11 Q9CX09	Q9cxg9 mus musculus
38	32	80.0	428	10 Q9LH89	Q9lh89 arabidopsis
39	32	80.0	571	10 Q24022	Q24022 lycopersico
40	32	80.0	901	12 Q99H09	Q99h09 helicoverpa
41	32	80.0	901	12 Q91BX6	Q91bx6 helicoverpa
42	32	80.0	978	5 Q9U6W5	Q9u6w5 caenorhabdi
43	32	80.0	1257	10 Q64516	Q64516 arabidopsis
44	31	77.5	42	2 Q9R730	Q9r730 xanthomonas
45	31	77.5	42	8 Q32948	Q32948 pinus thunb

ALIGNMENTS

RESULT 1

Q9CY64	PRELIMINARY;	PRT;	295 AA.
ID	Q9CY64		
AC	Q9CY64;		
DT	01-JUN-2001 (TREMREL. 17, Created)		
DT	01-JUN-2001 (TREMREL. 17, Last sequence update)		
DT	01-JUN-2001 (TREMREL. 17, Last annotation update)		
DE	2500001N03RIK PROTEIN.		
GN	2500001N03RIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,		
RA	Hayashizaki Y.;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-690(2001).		
DR	EMBL; AK010847; BAB2719.1; -.		
DR	MGI; 1917355; 2500001N03RIK.		

DR InterPro: IPR000683; GFO_IDH_MocA.
 DR Pfam: PF01408; GFO_IDH_MocA; 1.
 SQ SEQUENCE 295 AA; 33524 MW; F2E1682BD77032A4 CRC64;

Query Match 100.0%; Score 40; DB 11; Length 295;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||||

Db 274 KKRILHC 280

RESULT 2

Q9BRW8 PRELIMINARY; PRT; 296 AA.
 ID Q9BRW8
 AC Q9BRW8; 17, Created
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 19, Last annotation update)
 DE STIMUL TO BILIVERDIN REDUCTASE A.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN, PRIMITIVE NEUROECTODERMAL;
 RA Strausberg R.; 2001 to the EMBL/GenBank/DBJ databases.
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC005902; AAH05902.1; -
 DR InterPro: IPR000683; GFO_IDH_MocA.
 DR Pfam: PF01408; GFO_IDH_MocA; 1.
 SQ SEQUENCE 296 AA; 33432 MW; A389AFDDDB6ADBABF CRC64;

Query Match 100.0%; Score 40; DB 4; Length 296;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||||

Db 275 KKRILHC 281

RESULT 3

Q96QL4 PRELIMINARY; PRT; 296 AA.
 ID Q96QL4
 AC Q96QL4; 19, Created
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE UNKNOWN (PROTEIN FOR MGC:14706).
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PROSTATE;
 RA Strausberg R.; 2001 to the EMBL/GenBank/DBJ databases.
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC008456; AAH08456.1; -
 SQ SEQUENCE 296 AA; 33428 MW; 2CF2AA7F1CDDDB707 CRC64;

Query Match 100.0%; Score 40; DB 4; Length 296;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||||

Db 275 KKRILHC 281

RESULT 4

Q9DD21 PRELIMINARY; PRT; 303 AA.
 ID Q9DD21
 AC Q9DD21; 17, Created
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE 0610006A1LRK PROTEIN.
 GN 0610006A1LRK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK002231; BAB21950.1; -
 DR MGD; MGI:1915580; 0610006A1LRK.
 DR InterPro: IPR000683; GFO_IDH_MocA.
 DR Pfam: PF01408; GFO_IDH_MocA; 1.
 SQ SEQUENCE 303 AA; 34491 MW; 52D8A3B02BE956EB CRC64;

Query Match 100.0%; Score 40; DB 11; Length 303;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||||

Db 274 KKRILHC 280

RESULT 5

Q9LML0 PRELIMINARY; PRT; 512 AA.
 ID Q9LML0
 AC Q9LML0; 15, Created
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE F10K1.11 PROTEIN.
 GN F10K1.11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Liu S.X., Chan A., Yu G., Etgu P., Lee J.M., Lenz C., Pham P.,
 RA Sakano H., Toriumi M., Chung M., Goldsmith A., Liu A., Liu A.,
 RA Vaysberg M., Altati H., Brooks S., Buehler E., Chao Q., Conn L.,
 RA Conway A., Hansen M., Johnson-Hopson C., Khan S., Kim C., Lam B.,
 RA Miranda M., Nguyen M., Shinn P., Southwick A., Davis R.W.,

RA Ecker J.R., Federspiel N.A., Theologis A.;
 RT "The sequence of BAC F10K1 from Arabidopsis thaliana chromosome 1.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 DR EMBL: AC067971; AAF82202.1; -;
 DR InterPro: IPR001327; FAD pyr_redox.
 DR Pfam: PF00070; pyr_redox; 1.
 KW FAD; Flavoprotein; Oxidoreductase.
 SQ SEQUENCE 512 AA; 56857 MW; 1F63AFA9A1A2C13B CRC64;

Query Match 95.0%; Score 38; DB 10; Length 512;
 Best Local Similarity 85.7%; Pred. No. 4.4;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7
 III:III
 Db 236 KKRLLHC 242

RESULT 6
 Q9FHV3 PRELIMINARY; PRT; 599 AA.
 ID Q9FHV3
 AC Q9FHV3
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE MUTATOR-LIKE TRANSPOSABLE-LIKE PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=COLUMBIA;
 RC MEDLINE=99397451; PubMed=10470850;
 RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
 RT PI and TAC clones.";
 RL DNA Res. 6:183-195(1999).
 DR EMBL: AB017068; BAB11366.1; -;
 SQ SEQUENCE 599 AA; 69407 MW; 36BAEE2F2A82D717 CRC64;

Query Match 95.0%; Score 38; DB 10; Length 599;
 Best Local Similarity 85.7%; Pred. No. 5.1;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7
 III:III
 Db 563 KKRIMHC 569

RESULT 7
 Q49511 PRELIMINARY; PRT; 633 AA.
 ID Q49511
 AC Q49511
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE MUDR TRANSPOSABLE ELEMENT - LIKE PROTEIN (MUDR TRANSPOSABLE ELEMENT-LIKE PROTEIN).
 DE F28J12.70 OR AT4G18410.
 GN F28J12.70 OR AT4G18410.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RP Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Bancroft I., Mewes H.W., Mayer K., Schueller C.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL021710; CAI16721.1; -;
 DR EMBL: AL161548; CAB78843.1; -;
 DR InterPro: IPR004332; MUDR.
 DR Pfam: PF03108; MUDR; 1.
 SQ SEQUENCE 633 AA; 72930 MW; 3FC298BF2218C623 CRC64;

Query Match 95.0%; Score 38; DB 10; Length 633;
 Best Local Similarity 85.7%; Pred. No. 5.4;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7
 III:III
 Db 591 KKRIMHC 597

RESULT 8
 Q9QXV3 PRELIMINARY; PRT; 279 AA.
 ID Q9QXV3
 AC Q9QXV3; Q9QXV4; Q9QUP8; Q9QX3;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE INGI PROTEIN.
 GN INGI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=129/SVJ; TISSUE=BRAIN, EMBRYONIC FIBROBLAST, AND SPLEEN;
 RX MEDLINE=20011419; PubMed=10542254;
 RA Zerkenski M., Hill J.E., Kwek S.S.S., Grigorian I.A., Gurova K.V.,
 RA Garkavtsev I.V., Diatchenko L., Koonin E.V., Gudkov A.V.;
 RT "Structure and regulation of the mouse Ingi gene. Three alternative
 RT transcripts encode two PHD finger proteins that have opposite effects
 RT on p53 function.";
 RL J. Biol. Chem. 274:32172-32181(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Rancourt D., Garkavtsev I.;
 RT "Structural organization and expression pattern of the murine Ingi
 RT gene.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE-BREAST TUMOR;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ISOFORM 1 INHIBITS P53-DEPENDENT TRANSCRIPTIONAL
 CC ACTIVATION AND MAY FUNCTION AS AN ONCOPROTEIN. ISOFORM 2 ACTS AS A
 CC NEGATIVE GROWTH REGULATOR BY COOPERATING WITH P53 IN
 CC TRANSCRIPTIONAL ACTIVATION OF P53-RESPONSIVE GENES AND MAY ACT AS
 CC A TUMOR SUPPRESSOR.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: IN THE ADULT, WIDELY EXPRESSED WITH HIGHEST
 CC LEVELS IN THYMUS AND TESTIS. EXPRESSED THROUGHOUT THE WHOLE EMBRYO
 CC AT ALL STAGES OF DEVELOPMENT EXAMINED. AT DAY 10, HIGHEST
 CC EXPRESSION IS FOUND IN THE YOLK SAC WHILE AT DAY 16 AND 18, HIGHER
 CC LEVELS ARE FOUND IN INNER COMPARTMENTS OF BONE.
 CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST EXPRESSION OF ISOFORM

CC 1 IS FOUND AT DAY 11 WHILE HIGHEST EXPRESSION OF ISOFORM 2 IS
 CC FOUND AT DAY 7.

CC -1- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.

DR EMBL: AF177753; AAF16911.1; -

DR EMBL: AF177755; AAF16908.1; -

DR EMBL: AF177756; AAF16909.1; -

DR EMBL: AF177757; AAF16910.1; -

DR EMBL: AF149820; AAF09183.1; -

DR EMBL: BC016573; AAH16573.1; -

DR MGD: MGI:1349481; Ingi.

DR InterPro: IPR001965; PHD.

DR Pfam: PF00628; PHD; 1.

DR SMART: SM00249; PHD; 1.

KW Oncogene: Anti-oncogene; Alternative splicing.

FT DOMAIN 210 259 PHD-FINGER.

FT VARSPLIC 1 94 MISSING (IN ISOFORM 2).

FT CONFLICT 203 203 L -> F (IN REF. 2).

SQ SEQUENCE 279 AA; 33109 MW; 6765C984EEF179F4 CRC64;

Query Match 90.0%; Score 36; DB 11; Length 279;
 Best Local Similarity 71.4%; Pred. No. 6.8;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7

Db 67 KRRVLLHC 73

RESULT 9

Q9D6S2

ID Q9D6S2 PRELIMINARY; PRT; 508 AA.

AC Q9D6S2;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE 2-5' OLIGOADENYLATE SYNTHETASE-LIKE.

GN OASL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=TONGUE;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,

RA Hayashizaki Y.

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL: AK010034; BAB26655.1; -

DR MGD: MGI:1344390; Oasl.

DR InterPro: IPR001797; 25A-synth.

DR InterPro: IPR001201; PAP_25A.core.

DR FT CONFLICT 203 203 L -> F (IN REF. 2).

DR Pfam: PF00240; ubiqlutin; 1.

DR SMART: SM00213; UHQ; 1.

DR PROSITE: PS00833; 25A_SYNTH_2; UNKNOWN_1.

DR PROSITE: PSS0152; 25A_SYNTH_3; 1.

DR PROSITE: PSS0053; UBIQUITIN_2; 1.

SQ SEQUENCE 508 AA; 58767 MW; 9552B4540CC801A0 CRC64;

Query Match 90.0%; Score 36; DB 11; Length 508;

Best Local Similarity 71.4%; Pred. No. 12;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7

Db 109 KKRLLHC 115

RESULT 10

Q9AKS0

ID Q9AKS0 PRELIMINARY; PRT; 267 AA.

AC Q9AKS0;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE TATC PROTEIN.

GN TATC.

OS Pseudomonas stutzeri (Pseudomonas perfectomarina).

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=316;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ZOBELL ATCC14405;

RX MEDLINE=21101859; PubMed=11160097;

RA Heikkilae M.P., Honisch U., Wunsch P., Zumft W.G.;

RT "Role of the Tat transport system in nitrous oxide reductase

RT translocation and cytochrome cdi biosynthesis in Pseudomonas

RT stutzeri.";

RL J. Bacteriol. 183:1663-1671(2001).

DR EMBL: AJ299712; CAC29149.1; -

DR InterPro: IPR002033; UPF0032.

DR Pfam: PF00902; UPF0032; 1.

DR PROSITE: PS01218; UPF0032; 1.

SQ SEQUENCE 267 AA; 29439 MW; 4369E5E38BCD736 CRC64;

Query Match 87.5%; Score 35; DB 2; Length 267;

Best Local Similarity 71.4%; Pred. No. 11;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7

Db 21 KKRLLHC 27

RESULT 11

Q9ST63

ID Q9ST63 PRELIMINARY; PRT; 495 AA.

AC Q9ST63;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE PUTATIVE INTERNAL ROTENONE-INSENSITIVE NADH DEHYDROGENASE.

GN NDAL.

OS Solanum tuberosum (Potato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4113;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. DESIREE; TISSUE=LEAF;

RX MEDLINE=20040052; PubMed=10571867;

RA Rasmussen A.G., Svensson A., Knoop V., Grohmann L., Brennicke A.;

RT "Homologues of yeast and bacterial rotenone-insensitive NADH

RT dehydrogenases in higher eukaryotes: two enzymes are present in potato

RT mitochondria.";

RL Plant J. 20:79-87(1999).

CC -I- COFACTOR: FAD (BY SIMILARITY).
DR EMBL: AJ245861; CAB52796.1; -
DR InterPro: IPR001327; FAD_pyr_redox.
DR Pfam: PF00070; pyr_redox.1
KW FAD; Flavoprotein; NAD; Oxidoreductase.
SQ SEQUENCE 495 AA; 54902 MW; 6AFFC807BEB01340 CRC64;

Query Match 87.5%; Score 35; DB 10; Length 495;
Best Local Similarity 71.4%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7
|:|:|
Db 224 KRRLLHC 230

RESULT 12

Q941V1
ID Q941V1 PRELIMINARY; PRT; 499 AA.
AC Q941V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE NADH DEHYDROGENASE.
GN B1088C09.27.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
clone:B1088C09.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003734; BAB68119.1; -
SQ SEQUENCE 499 AA; 54687 MW; AF4337AB3F9363D4 CRC64;

Query Match 87.5%; Score 35; DB 10; Length 499;
Best Local Similarity 71.4%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7
|:|:|
Db 227 KRRLLHC 233

RESULT 13

O80874
ID O80874 PRELIMINARY; PRT; 508 AA.
AC O80874;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE NADH DEHYDROGENASE (UBIQUINONE OXIDOREDUCTASE).
GN AT2G29990.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=2083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
Buell C.R., Ketchum K.A., Lee J.-J., Rensing C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Mayam L., Tallon L.J., Gili J.E.,
Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -I- COFACTOR: FAD (BY SIMILARITY).
DR EMBL: AC004680; AAC31853.1; -
DR InterPro: IPR001327; FAD_pyr_redox.
DR Pfam: PF00070; pyr_redox.1.
KW FAD; Flavoprotein; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 508 AA; 56503 MW; 266A434E702A0C27 CRC64;

Query Match 87.5%; Score 35; DB 10; Length 508;
Best Local Similarity 71.4%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7
|:|:|
Db 236 KRRLLHC 242

RESULT 14

O9XZM5
ID O9XZM5 PRELIMINARY; PRT; 106 AA.
AC O9XZM5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RIBOSOMAL PROTEIN L44.
GN RL44.
OS Leishmania amazonensis.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5659;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lv78;
RA Porter-Kelley J., Chaudhuri G.;
RT "Cloning and characterization of Leishmania ribosomal protein L44.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE L44E FAMILY OF RIBOSOMAL PROTEINS.
DR EMBL: AF148853; AAD31928.2; -
DR InterPro: IPR000552; Ribosomal_L44E.
DR Pfam: PF00935; Ribosomal_L44; 1.
DR ProDom: PD002841; Ribosomal_L44E; 1.
DR ProSITE: PS01172; RIBOSOMAL_L44E; 1.
KW Ribosomal protein.
SQ SEQUENCE 106 AA; 12283 MW; F30A3AB2047B0334 CRC64;

Query Match 85.0%; Score 34; DB 5; Length 106;
Best Local Similarity 57.1%; Pred. No. 7.6;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7
|:|:|
Db 6 KKVVMHC 12

RESULT 15

O9H007
ID O9H007 PRELIMINARY; PRT; 111 AA.
AC O9H007;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE INGI ISOFORM (FRAGMENT).
GN INGI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20322670; PubMed=10866301;
RA Mehmet G., Ouchida M., Fukushima K., Hanafusa H., Etani T.,
RA Nishioka S., Nishizaki K., Shimizu K.;
RT "Genomic Structure of the Human INGI Gene and Tumor-specific Mutations
RT Detected in Head and Neck Squamous Cell Carcinomas.";
RL Cancer Res. 60:3143-3146(2000).
DR EMBL: AB037594; BAE20992.1; -.
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12566 MW; 9C76D21C7E982313 CRC64;

Query Match      85.0%; Score 34; DB 4; Length 111;
Best Local Similarity 71.4%; Pred. No. 7.9;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKRILHC 7
Db 50 KRRMLHC 56
```

Search completed: November 1, 2002, 14:49:50
Job time : 2.2717 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:41:32 ; Search time 0.360129 seconds
(without alignments)
752.611 Million cell updates/sec

Title: US-09-606-129A-34
Perfect score: 40
Sequence: 1 KKRILHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	% Query	Length	DB	ID	Description
1	40	100.0	296	1	BIEA_HUMAN	P53004	homo sapien
2	38	95.0	295	1	BIEA_RAT	P46844	rattus norv
3	36	90.0	473	1	OASL_MOUSE	Q92f2	mus musculus
4	33	82.5	528	1	DYR2_HUMAN	Q92630	homo sapien
5	32	80.0	571	1	AMPL_LYCES	Q10712	lycopersico
6	32	80.0	573	1	AMPL_SOLITU	P31427	solanum tub
7	32	80.0	925	1	PIP1_YEAST	P40020	saccharomyc
8	31	77.5	140	1	FKB2_MOUSE	P45878	mus musculus
9	31	77.5	141	1	FKB2_HUMAN	P26885	homo sapien
10	31	77.5	144	1	YHP5_YEAST	P38808	saccharomyc
11	31	77.5	257	1	Y2G1_CAEEL	P55326	caenorhabdi
12	31	77.5	423	1	SHP1_YEAST	P34223	saccharomyc
13	31	77.5	451	1	SUN_HAEIN	P44788	haemophilus
14	31	77.5	560	1	YD2H_SCHPO	Q10264	schizosacch
15	31	77.5	670	1	REP_HAEIN	P44804	haemophilus
16	31	77.5	1059	1	CAPU_DROME	Q24120	drosophila
17	30	75.0	108	1	Y209_METJA	O60271	methanococc
18	30	75.0	249	1	YAT6_RHOBL	P05449	rhodospseudo
19	30	75.0	323	1	CYCH_XENLA	P51947	xenopus lae
20	30	75.0	428	1	DGTP_MYCSM	O52199	mycobacteri
21	30	75.0	459	1	NEK3_HUMAN	P51956	homo sapien
22	30	75.0	588	1	DYR3_HUMAN	O43781	homo sapien
23	30	75.0	1790	1	SEPA_EMENI	P78621	emericaella
24	30	75.0	4128	1	PRKD_MOUSE	P97313	mus musculus
25	29	72.5	134	1	DC4C_ACICA	P20370	acinetobact
26	29	72.5	169	1	Y113_NPVAC	P41666	autographa
27	29	72.5	301	1	OTCA_VIBS2	P96172	vibriolo sp.
28	29	72.5	320	1	MDH_RHILV	O33525	rhizobium l
29	29	72.5	320	1	MDH_RHIME	O9eyj6	rhizobium m
30	29	72.5	328	1	HYBA_ECOLI	P37179	escherichia
31	29	72.5	359	1	ODPB_RAT	P49432	rattus norv
32	29	72.5	377	1	CYCI_HUMAN	Q14094	homo sapien
33	29	72.5	377	1	CYCI_MOUSE	Q922v9	mus musculus

RESULT 1				
BIEA_HUMAN	ID	BIEA_HUMAN	STANDARD;	PRT; 296 AA.
AC	P53004;			
DT	01-OCT-1996	(Rel. 34, Created)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Biliverdin reductase A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-			
DE	reductase).			
GN	BLVRA OR BLVR OR BVR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=96202961; PubMed=8631357;			
RA	Maines M.D., Polevoda B.V., Huang T.-J., McCoubrey W.K. Jr.;			
RT	"Human biliverdin IXalpha reductase is a zinc-metalloprotein.			
RT	Characterization of purified and Escherichia coli expressed			
RT	enzymes".			
RL	Eur. J. Biochem. 235:372-381(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Komuro A., Tobe T., Nakano Y., Yamaguchi T., Tomita M.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBDJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-117 FROM N.A.			
RA	Cordes M., Wollam C., Carter T.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBDJ databases.			
RN	[4]			
RP	SEQUENCE OF 3-36; 48-74 AND 228-248.			
RC	TISSUE=Liver;			
RX	MEDLINE=93143333; PubMed=8424666;			
RA	Maines M.D., Trakshel G.M.;			
RT	"Purification and characterization of human biliverdin reductase.";			
RL	Arch. Biochem. Biophys. 300:320-326(1993).			
RN	[5]			
RP	SEQUENCE OF 3-22.			
RC	TISSUE=Liver;			
RX	MEDLINE=95014177; PubMed=7929092;			
RA	Yamaguchi T., Komoda Y., Nakajima H.;			
RT	"Biliverdin-IX alpha reductase and biliverdin-IX beta reductase from			
RT	human liver. Purification and characterization.";			
RL	J. Biol. Chem. 269:24343-24348(1994).			
CC	J. FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN.			
CC	J. CATALYTIC ACTIVITY: BILIRUBIN + NAD(P)(+) = biliverdin + NAD(P)H.			
CC	COFACTOR: BINDS ONE ZINC ION. HAS DUAL PH/COFACTOR (NADH, NADPH)			
CC	SPECIFICITY. USES NADH AT THE ACIDIC PH RANGE (6-6.7) AND NADPH AT			
CC	THE ALKALINE RANGE (8.5-8.7).			
CC	J. PATHWAY: FINAL STEP IN HEME METABOLISM.			
CC	J. SUBUNIT: MONOMER.			
CC	J. SUBCELLULAR LOCATION: Cytoplasmic.			
CC	J. TISSUE SPECIFICITY: LIVER.			
CC	J. SIMILARITY: TO E.COLI YHHX.			

Q9x0x7 thermotoga
P25552 escherichia
P18140 tacaribe vi
P43704 haemophilus
O05542 gluconobact
P52468 human herpe
Q64347 mus musculu
P35524 rattus norv
O47077 enterobacte
O75643 homo sapien
P79114 bos taurus
Q9hd67 homo sapien

ALIGNMENTS

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X93086; CAA63635.1; --
DR EMBL: U34877; AAC35588.1; --
DR EMBL: AC005189; AAC25526.1; --
DR MIM: 109750; --
DR InterPro: IPR000683; GFO_IDH_MoCA.
DR Pfam: PF01408; GFO_IDH_MoCA; 1.
KW Oxidoreductase; NAD; NADP; Zinc.
FT PROPEP 1 2
FT CHAIN 3 296 BILIVERDIN REDUCTASE A.
FT DOMAIN 11 16 POLY-VAL.
FT METAL 280 280 ZINC (POTENTIAL).
FT METAL 281 281 ZINC (POTENTIAL).
FT METAL 292 292 ZINC (POTENTIAL).
FT METAL 293 293 ZINC (POTENTIAL).
FT METAL 293 293 ZINC (POTENTIAL).
FT CONFLICT 3 3 A -> T (IN REF. 2).
FT CONFLICT 154 155 SD -> AG (IN REF. 2).
FT CONFLICT 160 160 D -> E (IN REF. 2).
SQ SEQUENCE 296 AA; 33488 MW; 0DFD3B386F4DFC0A CRC64;

Query Match 100.0%; Score 40; DB 1; Length 296;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|||||||
Db 275 KKRILHC 281

RESULT 2
BIEA_RAT
ID BIEA_RAT STANDARD; PRT; 295 AA.
AC P46844;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Biliverdin reductase A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-
DE reductase).
GN BLVRA OR BLVR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=92156147; PubMed=1371282;
RA Fakhrai H., Maines M.D.;
RT "Expression and characterization of a cDNA for rat kidney biliverdin
RT reductase. Evidence suggesting the liver and kidney enzymes are the
RT same transcript product.";
RL J. Biol. Chem. 267:4023-4029(1992).
RN [2]
RP MUTAGENESIS.
RX MEDLINE=94291657; PubMed=8020496;
RA McCoubrey W.K. Jr., Maines M.D.;
RT "Site-directed mutagenesis of cysteine residues in biliverdin
RT reductase. Roles in substrate and cofactor binding.";
RL Eur. J. Biochem. 222:597-603(1994).
CC -1- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN; DISPLAYS TWO DISTINCT
CC PH OPTIMA USING A DIFFERENT COFACTOR AT EACH PH: NADH AT THE LOWER
CC PH 6.7-6.9 RANGE AND NADPH AT PH 8.5-8.7. NADPH, HOWEVER, IS THE
CC PROBABLE COFACTOR IN BIOLOGICAL SYSTEMS.
CC -1- CATALYTIC ACTIVITY: Bilirubin + NAD(P)(+) = biliverdin + NAD(P)H.
CC -1- COFACTOR: BINDS ONE ZINC ION.

CC -1- PATHWAY: FINAL STEP IN HEME METABOLISM.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: TO E.COLI YHHX.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M81681; AAA40830.1; --
DR InterPro: IPR000683; GFO_IDH_MoCA.
DR Pfam: PF01408; GFO_IDH_MoCA; 1.
KW Oxidoreductase; NAD; NADP; Zinc.
FT PROPEP 1 2
FT CHAIN 3 295 BILIVERDIN REDUCTASE A.
FT DOMAIN 11 16 POLY-VAL.
FT METAL 279 279 ZINC (POTENTIAL).
FT METAL 280 280 ZINC (POTENTIAL).
FT METAL 291 291 ZINC (POTENTIAL).
FT METAL 292 292 ZINC (POTENTIAL).
FT METAL 293 293 ZINC (POTENTIAL).
FT MUTAGEN 73 73 C->A: LOSS OF ACTIVITY.
FT MUTAGEN 280 280 C->A: REDUCED ACTIVITY.
FT MUTAGEN 291 291 C->A: REDUCED ACTIVITY.
SQ SEQUENCE 295 AA; 33565 MW; 219C8EA96C150588 CRC64;

Query Match 95.0%; Score 38; DB 1; Length 295;
Best Local Similarity 85.7%; Pred. No. 0.77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|||||
Db 274 KKRIMHC 280

RESULT 3
OASL_MOUSE
ID OASL_MOUSE STANDARD; PRT; 473 AA.
AC Q9Z2F2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 54 kDa 2'-5'-oligoadenylate synthetase like protein (EC 2.7.7.-) (p54
DE OASL) (P54OASL) (M1204).
GN OASL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=99323699; PubMed=10395668;
RA Tiefenthaler M., Marksteiner R., Neyer S., Koch F., Hofer S.,
RA Schuler G., Nussenzweig M., Schneider R., Heufler C.,
RT "M1204, a novel 2',5' oligoadenylate synthetase with a ubiquitin-like
RT extension, is induced during maturation of murine dendritic cells.";
RL J. Immunol. 163:760-765(1999).
CC -1- FUNCTION: MAY PLAY A ROLE IN MEDIATING RESISTANCE TO VIRUS
CC INFECTION, CONTROL OF CELL GROWTH, DIFFERENTIATION, AND APOPTOSIS.
CC -1- CATALYTIC ACTIVITY: BINDS DOUBLE-STRANDED RNA AND POLYMERIZES ATP
CC INTO PP(A2'P5'A)N OLIGOMERS, WHICH ACTIVATE THE LATENT RNASE L
CC THAT, WHEN ACTIVATED, CLEAVES SINGLE-STRANDED RNAs.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SPLEEN DENDRITIC CELLS,
CC WHEREAS, IN BONE MARROW-DERIVED DENDRITIC CELLS, THE AMOUNT
CC INCREASES DURING THE MATURATION PROCESS. EXPRESSED IN MANY ORGANS,
CC THE HIGHEST LEVELS BEING IN THYMUS, LUNG, AND BONE MARROW.
CC -1- SIMILARITY: BELONGS TO THE 2-5A SYNTHETASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 UBIQUITIN-LIKE DOMAIN.
CC -1- CAUTION: THIS MAY NOT BE THE TRUE ORTHOLOG OF HUMAN OASL.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF068835; AAD02818.1; -
MGD; MGI:1344390; Oasi.
InterPro; IPR001797; 25A_synth.
InterPro; IPR001201; PAP_25A_Core.
InterPro; IPR000626; ubiquitin.
Pfam; PF00240; ubiquitin; 1.
PROSITE; PS00832; 25A_SYNTH_1; FALSE_NEG.
PROSITE; PS00833; 25A_SYNTH_2; 1.
PROSITE; PS0152; 25A_SYNTH_3; 1.
PROSITE; PS50053; UBIQUITIN_2; FALSE_NEG.
RNA-binding; Transferase; Nucleotidyltransferase.
DOMAIN 435 473 UBIQUITIN-LIKE.
SEQUENCE 473 AA: 54625 MW: 57060E08A51C8460 CRC64;

Query Match 90.08; Score 36; DB 1; Length 473;
Best Local Similarity 71.48; Pred. No. 3.2;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7
| | | | |
Db 109 KKRILHC 115

RESULT 4
ID Dyr2_HUMAN STANDARD; PRT; 528 AA.
AC Q92630;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 2
DE (EC 2.7.1.-).
GN DYRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98421512; PubMed=9748265;
RA Becker W., Weber Y., Wetzel K., Eirmbter K., Tejedor F.J., Joost H.-G.;
RA "Sequence characteristics, subcellular localization, and substrate specificity of DYRK-related kinases, a novel family of dual specificity protein kinases."
RL J. Biol. Chem. 273:25893-25902(1998).
RN [2]
RP SEQUENCE OF 320-528 FROM N.A.
RC TISSUE=Placenta;
RA Becker W., Joost H.-G.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN VITRO: CAN PHOSPHORYLATE HISTONES H3 AND H2B ON SER AND THR RESIDUES. MAY BE INVOLVED IN THE REGULATION OF CELLULAR GROWTH AND/OR DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. MNB/DYRK SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Y13493; CAA73885.1; -
EMBL; Y09216; CAA70418.1; -
HSP; Q16539; 1WFC.
MIM: 603496; -
InterPro; IPR000719; Euk_pkinase
InterPro; IPR002290; Ser_thr_pkinase.
Pfam; PF00069; pkinase; 2.
SMART; SM00220; S_TKC; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase.
KW ATP-binding; Phosphorylation.
DOMAIN 149 462 PROTEIN KINASE.
FT NP_BIND 155 163 ATP (BY SIMILARITY).
FT BINDING 178 178 ATP (BY SIMILARITY).
FT ACT_SITE 275 275 BY SIMILARITY.
SEQUENCE 528 AA: 59714 MW: AF2C8622ED9522D7 CRC64;

Query Match 82.58; Score 33; DB 1; Length 528;
Best Local Similarity 71.48; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRILHC 7
| | | | |
Db 268 KKRILHC 274

RESULT 5
ID AMP1_LYCES STANDARD; PRT; 571 AA.
AC Q10712; Q9S9A3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Amino-peptidase 1, chloroplast precursor (EC 3.4.11.1) (Leucine amino-peptidase) (LAP) (leucyl aminopeptidase) (proline aminopeptidase) (EC 3.4.11.5) (Prolyl aminopeptidase) (DR57).
GN LAP1 OR LAP OR LAP2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Peto 238R; TISSUE=Leaf;
RX MEDLINE=96421572; PubMed=8824220;
RA Gu Y.Q., Chao W.S., Walling L.L.;
RA "Localization and post-translational processing of the wound-induced leucine aminopeptidase proteins of tomato."
RL J. Biol. Chem. 271:25880-25887(1996).
RN [2]
RP SEQUENCE OF 49-571 FROM N.A.
RC STRAIN=CV. VF36; TISSUE=Pistill;
RX MEDLINE=95375233; PubMed=7647301;
RA Milligan S.B., Gasser C.S.;
RA "Nature and regulation of pistill-expressed genes in tomato."
RL Plant Mol. Biol. 28:691-711(1995).
RN [3]
RP SEQUENCE OF 103-571 FROM N.A.
RC STRAIN=CV. Peto 238R; TISSUE=Leaf;
RX MEDLINE=94052201; PubMed=8234334;
RA Pautot V., Holzer F.M., Reisch B., Walling L.L.;
RA "Leucine aminopeptidase: an inducible component of the defense response in Lycopersicon esculentum (tomato)."
RL Proc. Natl. Acad. Sci. U.S.A. 90:9906-9910(1993).
CC -1- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR TURNOVER OF INTRACELLULAR PROTEINS.

CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids
CC including Pro although not Arg or Lys, and Xbb may be Pro.
CC -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- INDUCTION: BY WOUNDING.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17; ALSO KNOWN AS THE
CC CYTOSOL AMINOPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U50151; AAC49456.1; -;
CC EMBL; U50152; AAC49457.1; -;
CC EMBL; U20593; AAA80498.1; -;
CC HSSP; P00727; ILAM.
CC MEROPS; M17.002; -;
CC InterPro: IPR000819; Peptidase_M17.
CC Pfam; PF00883; Peptidase_M17; 1.
CC Pfam; PF02789; Peptidase_M17_N; 1.
CC PRINTS; PR00481; LAMNOPPTDASE.
CC PROSITE; PS00631; CYTOSOLAP; 1.
CC TRANSIT peptide; Chloroplast; Amino-peptidase; Hydrolase; Zinc.
CC TRANSIT 1 53 CHLOROPLAST (POTENTIAL).
CC CHAIN 54 571 AMINOPEPTIDASE 1.
CC DOMAIN 169 174 POLY-ALA.
CC METAL 342 342 ZINC 2 (BY SIMILARITY).
CC METAL 347 347 ZINC 1 AND 2 (BY SIMILARITY).
CC METAL 367 367 ZINC 2 (BY SIMILARITY).
CC METAL 427 427 ZINC 1 (BY SIMILARITY).
CC METAL 429 429 ZINC 1 AND 2 (BY SIMILARITY).
CC ACT_SITE 354 354 POTENTIAL.
CC ACT_SITE 431 431 POTENTIAL.
CC VARIANT 358 358 R -> G.
CC CONFLICT 271 271 P -> N (IN REF. 2).
CC CONFLICT 315 315 T -> S (IN REF. 3).
CC CONFLICT 515 515 T -> L (IN CLONE PBLAP2).
CC SEQUENCE 571 AA; 60279 MW; C7A224837E73939D CRC64;
Query Match 80.0%; Score 32; DB 1; Length 571;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KRILHC 7 PRT; 573 AA.
Db 48 KRIVHC 53
RESULT 6
ID AMPL_SOLUTU STANDARD; PRT; 573 AA.
AC P31427;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Amino-peptidase, chloroplast precursor (EC 3.4.11.1) (Leucine
DE amino-peptidase, chloroplast precursor (EC 3.4.11.1) (Leucine
DE amino-peptidase) (LAP) (Leucyl amino-peptidase) (Proline amino-peptidase)
DE (EC 3.4.11.5) (Prolyl amino-peptidase).
GN LAP.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. DESIREE;
RX MEDLINE=94339796; PubMed=7765119;

RA Herbers K., Prat S., Willmitzer L.;
RT "Functional analysis of a leucine aminopeptidase from Solanum
RL tuberosum L.";
RN Planta 194:230-240(1994).
[2]
RP SEQUENCE OF 19-573 FROM N.A.
RC STRAIN=CV. DESIREE; TISSUE=Leaf;
RX MEDLINE=93005746; PubMed=1392612;
RA Hildmann T., Ebner M., Pena-Cortes H., Sanchez-Serrano J.J.,
RA Willmitzer L., Prat S.;
RT "General roles of abscisic and jasmonic acids in gene activation as a
RT result of mechanical wounding.";
RL Plant Cell 4:1157-1170(1992).
CC -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
CC TURNOVER OF INTRACELLULAR PROTEINS.
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids
CC including Pro although not Arg or Lys, and Xbb may be Pro.
CC -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC -!- SUBUNIT: HOMOHETEROMER (PROBABLE).
CC -!- SUBCELLULAR LOCATION: Chloroplast (By similarity).
CC -!- TISSUE SPECIFICITY: IN TUBERS AND FLORAL BUDS OF UNTREATED PLANTS.
CC AFTER ABA TREATMENT OR MECHANICAL WOUNDING IS MOSTLY ACCUMULATED
CC IN LEAVES, TO A LESSER EXTENT IN STEMS, BUT NOT IN ROOTS.
CC -!- INDUCTION: BY ABSICISIC ACID (ABA), JASMONIC ACID (JA) AND
CC WOUNDING.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17; ALSO KNOWN AS THE
CC CYTOSOL AMINOPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X77015; CAA54314.1; -;
CC EMBL; X67845; CAA48038.1; -;
CC PIR; S24769; S24769.
CC PIR; PQ0470; PQ0470.
CC HSSP; P00727; ILAM.
CC MEROPS; M17.002; -;
CC InterPro: IPR000819; Peptidase_M17.
CC Pfam; PF00883; Peptidase_M17; 1.
CC Pfam; PF02789; Peptidase_M17_N; 1.
CC PRINTS; PR00481; LAMNOPPTDASE.
CC PROSITE; PS00631; CYTOSOLAP; 1.
CC TRANSIT peptide; Chloroplast; Amino-peptidase; Hydrolase; Zinc.
CC TRANSIT 1 53 CHLOROPLAST (POTENTIAL).
CC CHAIN 54 573 AMINOPEPTIDASE.
CC DOMAIN 169 174 POLY-ALA.
CC METAL 342 342 ZINC 2 (BY SIMILARITY).
CC METAL 347 347 ZINC 1 AND 2 (BY SIMILARITY).
CC METAL 367 367 ZINC 2 (BY SIMILARITY).
CC METAL 427 427 ZINC 1 (BY SIMILARITY).
CC METAL 429 429 ZINC 1 AND 2 (BY SIMILARITY).
CC ACT_SITE 354 354 POTENTIAL.
CC ACT_SITE 431 431 POTENTIAL.
CC SEQUENCE 573 AA; 60122 MW; 3152145A47FB291 CRC64;
Query Match 80.0%; Score 32; DB 1; Length 573;
Best Local Similarity 83.3%; Pred. No. 27;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KRILHC 7
Db 48 KRIVHC 53
RESULT 7
ID PIP1_YEAST STANDARD; PRT; 925 AA.
PIPI_YEAST

AC P40020;
 DT 01-NOV-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Polymerase-interacting protein 1 (factor interacting with REF).
 GN PIP1 OR FIR1 OR YER032W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W303;
 RA del Olmo M., Gross S., Moore C.L.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Aviles E., Berio A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RA MEDLINE=97339480; PubMed=9196079;
 RX Russnak R., Pereira S., Platt T.;
 RT "RNA binding analysis of yeast REF2 and its two-hybrid interaction
 RL with a new gene product, FIR1.";
 RL Gene Expr. 6:241-258(1996).
 CC -1- FUNCTION: Interacts with poly(A) polymerase and with REF2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U17262; AAB46625.1; -;
 DR EMBL; U18778; AAB64565.1; -;
 DR SCD; S0000834; FIR1.
 FT CONFLICT 663 663 R -> P (IN REF. 2).
 SQ SEQUENCE 925 AA; 104701 MW; 707D9839EE31322B CRC64;
 Query Match 80.0%; Score 32; DB 1; Length 925;
 Best Local Similarity 71.4%; Pred. No. 43;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 KKRILHC 7
 Db 841 KKRILHC 847
 RESULT 8
 FKB2_MOUSE
 ID FKB2_MOUSE STANDARD; PRT; 140 AA.
 AC P45878;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE FK506-binding protein precursor (FKBP-13) (peptidyl-prolyl cis-trans
 DE isomerase) (PPIase) (EC 5.2.1.8).
 GN FKB2 OR FKBP13.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=129/SVJ; TISSUE=Liver;
 RX MEDLINE=94085790; PubMed=7505249;
 RA Hendrickson B.A., Zhang W., Craig R.J., Jin Y.J., Bierer R.E.,
 RA Burakoff S.J., Dilella A.G.;
 RT "Structural organization of the genes encoding human and murine
 RT FK506-binding protein (FKBP) 13 and comparison to FKBP1.";
 RL Gene 134:271-275(1993).
 CC -1- FUNCTION: PPIases accelerate the folding of proteins. May function
 CC as a component of membrane cytoskeletal scaffolds.
 CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
 CC PEPTIDE BONDS IN OLIGOPEPTIDES.
 CC -1- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
 CC -1- SUBUNIT: Interacts with the C-terminal domain of 4.1G.
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN. MEMBRANE
 CC ASSOCIATED (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M77831; AAA37631.1; -;
 DR HSP; P20081; IYAT.
 DR MGD; MGI:95342; FKBP2.
 DR InterPro; IPR001179; FKBP_PPIase.
 DR Pfam; PF00254; FKBP; 1.
 DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
 DR PROSITE; PS00454; FKBP_PPIASE_2; 1.
 DR PROSITE; PS00059; FKBP_PPIASE_3; 1.
 KW Isomerase; Rotamase; Signal; Endoplasmic reticulum.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 140 FK506-BINDING PROTEIN.
 FT SITE 137 140 PREVENT SECRETION FROM ER (POTENTIAL).
 SQ SEQUENCE 140 AA; 15344 MW; F4E7FCC7766A0416 CRC64;
 Query Match 77.5%; Score 31; DB 1; Length 140;
 Best Local Similarity 71.4%; Pred. No. 11;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 KKRILHC 7
 Db 34 KKRVDHC 40
 RESULT 9
 FKB2_HUMAN
 ID FKB2_HUMAN STANDARD; PRT; 141 AA.
 AC P26885;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE FK506-binding protein precursor (FKBP-13) (peptidyl-prolyl cis-trans
 DE isomerase) (PPIase) (EC 5.2.1.8).
 GN FKB2 OR FKBP13.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Colon carcinoma;
 RX MEDLINE=91319747; PubMed=1713687;
 RA Jin Y.-J., Albers M.W., Lane W.S., Bierer B.E., Schreiber S.L.,
 RA Burakoff S.J.;
 RT "Molecular cloning of a membrane-associated human FK506- and
 RT rapamycin-binding protein, FKBP-13.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6677-6681(1991).
 RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=93112052; PubMed=1281998;
RA Dillella A.G., Hawkins A., Craig R.J., Schreiber S.L., Griffin C.A.;
RT "Chromosomal band assignments of the genes encoding human FKBP12 and
FKBP13".
RL Biochem. Biophys. Res. Commun. 189:819-823(1992).
CC -1- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN. MEMBRANE
ASSOCIATED (PROBABLE).
CC -1- TISSUE SPECIFICITY: T-CELLS AND THYMUS.
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M65128; AAA58473.1; -;
CC EMBL; M75099; AAA36563.1; -;
CC PIR; JCI365; JC1365.
CC HSP; Q00686; lpbk.
CC MIM; 186946; -;
CC InterPro: IPR001179; FKBP_PPIase.
CC Pfam; PF00254; FKBP; 1.
CC PROSITE; PS00453; FKBP_PPIASE_1; 1.
CC PROSITE; PS00454; FKBP_PPIASE_2; 1.
CC PROSITE; PS00059; FKBP_PPIASE_3; 1.
CC Isomerase; Rotamase; Signal; Endoplasmic reticulum; Polymorphism.
KW SIGNAL 1 21
FT CHAIN 22 141 FK506-BINDING PROTEIN.
FT SITE 138 141 PREVENT SECRETION FROM ER (POTENTIAL).
FT VARIANT 21 21 S -> TA.
FT VARIANT 24 24 /FTid=VAR_006410.
FT VARIANT 96 96 T -> A.
FT VARIANT 96 96 /FTid=VAR_006411.
FT VARIANT 96 96 /FTid=VAR_006412.
FT SEQUENCE 141 AA; 15654 MW; 9F4751CA7D82D064 CRC64;
Query Match 77.5%; Score 31; DB 1; Length 141;
Best Local Similarity 71.4%; Pred. No. 11;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KKRILHC 7
III; II
Db 35 KKRVDHC 41
RESULT 10
YHP5_YEAST STANDARD; PRT; 144 AA.
AC P38808;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 16.7 kDa protein in HXT1-HXT5 intergenic region.
GN YHR095W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,
N.

RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII".
RL Science 265:2077-2082(1994).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U00060; AAB69924.1; -;
CC PIR; S46716; S46716.
CC SGD; S0001137; YHR095W.
KW Hypothetical protein.
KW SEQUENCE 144 AA; 16678 MW; 1AB89491523EEA89 CRC64;
Query Match 77.5%; Score 31; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKRILH 6
IIIIII
Db 118 KKRILH 123
RESULT 11
YZG1_CAEL STANDARD; PRT; 257 AA.
ID YZG1_CAEL
AC P55326; Q19405;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 28.7 kDa protein F13E6.1 in chromosome X.
GN F13E6.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Margerison S.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE TPD52 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z68105; CAA92122.1; -;
CC WormPep; F13E6.1; CE03198.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 48 97 COILED COIL (POTENTIAL).
FT SEQUENCE 257 AA; 28696 MW; 2A2A667672A8328A CRC64;
Query Match 77.5%; Score 31; DB 1; Length 257;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKRILH 6
IIIIII
Db 157 KKRILH 162

RA Pearson D., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO YEAST SNM1
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Z99295; CAB16571.2; -;
DR EMBL; Z69728; CAA93588.1; -;
KW Hypothetical protein.
SQ SEQUENCE 560 AA; 63076 MW; BF4BDE8C28EAD227 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 560;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRILHC 7
Db 295 RRVLHC 300
:|:|:|

RESULT 15
REP_HAEIN STANDARD; PRT; 670 AA.
ID REP_HAEIN STANDARD; PRT; 670 AA.
AC P44804;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent DNA helicase rep (EC 3.6.1.-).
GN REP OR HI0649.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: REP HELICASE IS A SINGLE-STRANDED DNA-DEPENDENT
CC ATPASE INVOLVED IN DNA REPLICATION; IT CAN INITIATE UNWINDING
CC AT A NICK IN THE DNA. IT BINDS TO THE SINGLE-STRANDED DNA AND
CC ACTS IN A PROGRESSIVE FASHION ALONG THE DNA IN THE 3' TO 5'
CC DIRECTION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U32748; AAC22309.1; -;
DR HSSP; P09980; 1UAA.
TIGR; HI0649; -;

DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
KW Hydrolase; Helicase; DNA replication; ATP-binding; DNA-binding;
KW Complete proteome.
FT NP_BIND 22 29 ATP (POTENTIAL).
SQ SEQUENCE 670 AA; 77745 MW; 802A131251ED4E24 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 670;
Best Local Similarity 83.3%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRILHC 7
Db 281 QRILHC 286
:|:|:|

Search completed: November 1, 2002, 14:47:49
Job time : 1.36013 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:58 ; Search time 0.742765 Seconds
(without alignments)
905.569 Million cell updates/sec

Title: US-09-606-129a-34
Perfect score: 40
Sequence: 1 KKRILHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	296	G02066	billiverdin reducta
2	40	100.0	296	S62624	billiverdin reducta
3	38	95.0	295	A42268	billiverdin reducta
4	38	95.0	512	H86206	hypothetical prote
5	38	95.0	633	T04537	hypothetical prote
6	35	87.5	508	T02486	hypothetical prote
7	34	85.0	457	C86463	hypothetical prote
8	33	82.5	137	T46904	hypothetical prote
9	33	82.5	401	F85015	probable phosphati
10	33	82.5	431	T01723	1-phosphatidylinos
11	32	80.0	144	H41700	C6 protein - rabbi
12	32	80.0	189	A38117	hypothetical prote
13	32	80.0	189	AD2414	transposase all486
14	32	80.0	189	AE2064	transposase all206
15	32	80.0	189	AF2052	transposase all197
16	32	80.0	189	AF2152	transposase alr277
17	32	80.0	189	AG1875	transposase alr055
18	32	80.0	189	AG2037	transposase alr185
19	32	80.0	383	S51651	cyclin delta-2 - A
20	32	80.0	566	T07850	leucyl aminopeptid
21	32	80.0	571	T07047	leucyl aminopeptid
22	32	80.0	571	T07849	leucyl aminopeptid
23	32	80.0	573	S41376	leucyl aminopeptid
24	32	80.0	925	S50490	hypothetical prote
25	32	80.0	1257	T01020	hypothetical prote
26	31	77.5	42	T07474	hypothetical prote
27	31	77.5	100	AI0092	probable membrane
28	31	77.5	140	I49668	binding protein -
29	31	77.5	142	JC1365	FK506/rapamycin-bi

30	31	77.5	144	2	S46716	hypothetical prote
31	31	77.5	145	2	F75189	hypothetical prote
32	31	77.5	257	2	E89632	protein F1366.1 li
33	31	77.5	423	2	S39830	hypothetical prote
34	31	77.5	451	1	F64155	hypothetical prote
35	31	77.5	502	2	T40792	hypothetical prote
36	31	77.5	504	2	JC4775	interferon-induced
37	31	77.5	504	2	A56534	hypothetical prote
38	31	77.5	508	2	T22440	hypothetical prote
39	31	77.5	522	2	E86453	CDS protein F911.
40	31	77.5	539	2	T38927	hypothetical prote
41	31	77.5	551	2	T11629	NADH dehydrogenase
42	31	77.5	698	2	D64084	helicase (EC 3.6.1
43	31	77.5	707	2	T34444	hypothetical prote
44	31	77.5	817	2	T22442	hypothetical prote
45	31	77.5	1058	2	T13286	cappuccino gene pr

ALIGNMENTS

RESULT 1

G02066
billiverdin reductase (EC 1.3.1.24) - human
N:Alternate names: billiverdin IX-alpha reductase
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G02066
R:Komuro, A.; Tobe, T.; Nakano, Y.; Yamaguchi, T.; Tomita, M.
submitted to the EMBL Data Library, August 1995
A:Reference number: H00768
A:Accession: G02066
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-296 <COM>
C:Cross-references: EMBL:U34877; NID:g1143231; PID:g1143232
C:Keywords: oxidoreductase

Query Match 100.0%; Score 40; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|||||||
DB 275 KKRILHC 281

RESULT 2

S62624
billiverdin reductase (EC 1.3.1.24) - human
N:Alternate names: billiverdin IX-alpha reductase
C:Species: Homo sapiens (man)
C>Date: 28-Oct-1996 #sequence_revision 09-May-1997 #text_change 21-Jul-2000
C:Accession: S62624; S62622; S29736
R:Maines, M.D.; Polevoda, B.V.; Huang, T.J.; McCoubrey Jr., W.K.
Eur. J. Biochem. 235, 372-381, 1996
A:Title: Human billiverdin IX-alpha reductase is a zinc-metalloprotein. Characterization
A:Reference number: S62622; MUID:96202961
A:Accession: S62624
A:Molecule type: mRNA
A:Residues: 1-296 <MAI>
A:Cross-references: EMBL:X93086; NID:g1246748; PIDN:CAA63635.1; PID:g1246749
A:Accession: S62622
A:Molecule type: protein
A:Residues: 3-24,'X',26-27,'X',29-36;48-74;228-234;235-248 <NAF>
R:Maines, M.D.; Trakshel, G.M.
Arch. Biochem. Biophys. 300, 320-326, 1993
A:Title: Purification and characterization of human billiverdin reductase.
A:Reference number: S29736; MUID:93143333
A:Accession: S29736
A:Molecule type: protein
A:Residues: 3-24,'X',26-27,'X',29-36;48-74;228-234;235-248 <NAW>
A:Note: the sequence of peptide 1 from page 323 seems not to belong to this protein

C;Genetics:

A;Gene: BVR

C;Keywords: oxidoreductase

F;3-296/Product: biliverdin reductase IX-alpha #status experimental <MAT>

Query Match 100.0%; Score 40; DB 2; Length 296;

Best Local Similarity 100.0%; Pred. No. 0.96;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||

Db 275 KKRILHC 281

RESULT 3

A42268

biliverdin reductase (EC 1.3.1.24) - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C;Accession: A42268

R;Fakhrai, H.; Maines, M.D.

J. Biol. Chem. 267, 4023-4029, 1992

A;Title: Expression and characterization of a cDNA for rat kidney biliverdin reductase.

A;Reference number: A42268; MUID:92156147

A;Accession: A42268

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: nucleic acid; protein

A;Residues: 1-295 <FAK>

A;Cross-references: GB:M81681; NID:g203177; PIDN:AAA40830.1; PID:g203178

A;Experimental source: kidney

A;Note: sequence extracted from NCBI backbone (NCBIP:82800)

C;Keywords: liver; oxidoreductase

Query Match

Best Local Similarity 95.0%; Score 38; DB 2; Length 295;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||

Db 274 KKRIMHC 280

RESULT 4

H86206

hypothetical protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C;Accession: H86206

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719

A;Accession: H86206

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-512 <STO>

A;Cross-references: GB:AE005172; NID:g8954028; PIDN:AAF82202.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

Query Match

Best Local Similarity 95.0%; Score 38; DB 2; Length 512;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||

Db 236 KKRLLHC 242

RESULT 5

T04537

hypothetical protein F28J12.70 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999

C;Accession: T04537

R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro

submitted to the Protein Sequence Database, February 1998

A;Reference number: Z15377

A;Accession: T04537

A;Molecule type: DNA

A;Residues: 1-633 <BEV>

A;Cross-references: EMBL:AL021710

A;Experimental source: cultivar Columbia; BAC clone F28J12

C;Genetics:

A;Map position: 4

A;Introns: 281/3; 303/3; 442/1; 614/3

A;Note: F28J12.70

Query Match

Best Local Similarity 95.0%; Score 38; DB 2; Length 633;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||

Db 591 KKRIMHC 597

RESULT 6

T02486

hypothetical protein At2g29990 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein F23F1.9

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001

C;Accession: T02486; B84703

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K

submitted to the EMBL Data Library, August 1998

A;Description: Arabidopsis thaliana chromosome II BAC F23F1 genomic sequence.

A;Reference number: Z14675

A;Accession: T02486

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-508 <ROU>

A;Cross-references: EMBL:AC004680; NID:g3420043; PID:g3420052

A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487

A;Accession: B84703

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-508 <STO>

A;Cross-references: GB:AE002093; NID:g3420052; PIDN:AAC31853.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g29990; F23F1.9

A;Map position: 2

A;Introns: 158/3; 230/1; 283/3; 305/3; 360/2; 398/3; 458/3

C;Superfamily: NADH dehydrogenase

Query Match

Best Local Similarity 87.5%; Score 35; DB 2; Length 508;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|||||

Db 236 KKRLLHC 242

```

RESULT 7
C86463
hypothetical protein T3M13.4 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: C86463
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chln, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86463
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-497 <SPO>
A:Cross-references: GB:AE005172; NID:g6957509; PIDN:AAF32431.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match      85.0%; Score 34; DB 2; Length 497;
Best Local Similarity 83.3%; Pred. No. 27;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KRILHC 7
    |||||
Db 434 KRVLHC 439

RESULT 8
T46904
hypothetical protein DKFzp761D081.1 - human
C:Species: Homo sapiens (man)
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 02-Sep-2000
C:Accession: T46904
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24136
A:Accession: T46904
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-137 <AAA>
A:Cross-references: EMBL:AL157474
A:Experimental source: adult amygdala; clone DKFzp761D081
C:Genetics:
A:Note: DKFzp761D081.1
A:Superfamily: RING finger homology
F:81-131/Domain: RING finger homology <RRN>

Query Match      82.5%; Score 33; DB 2; Length 137;
Best Local Similarity 85.7%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRILHC 7
    ||| |||
Db 63 KKRKLHC 69

RESULT 9
F85015
probable phosphatidylinositol kinase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: F85015
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

```

```

A:Reference number: A85001; MUID:20083488
A:Accession: F85015
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <SPO>
A:Cross-references: GB:NC_001268; NID:g7267616; PIDN:CAB80928.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g01190
A:Map position: 4

```

```

Query Match      82.5%; Score 33; DB 2; Length 401;
Best Local Similarity 85.7%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 KKRILHC 7
    ||||| ||
Db 353 KRRIHC 359

```

```

RESULT 10
T01723

```

```

1-phosphatidylinositol-4-phosphate 5-kinase type II homolog - Arabidopsis thaliana
N:Alternate names: protein A_IG002N01.9
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01723
R:Scheet, P.; Maggil, L.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana IG002N01.
A:Reference number: Z14407
A:Accession: T01723
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-431 <SCH>
A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191143
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 40/2; 94/3; 161/3; 224/2; 255/1; 271/1; 303/1; 339/2
A:Note: A_IG002N01.9

```

```

Query Match      82.5%; Score 33; DB 2; Length 431;
Best Local Similarity 85.7%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 KKRILHC 7
    ||||| ||
Db 383 KRRIHC 389

```

```

RESULT 11
H41700

```

```

C6 protein - rabbit fibroma virus
C:Species: rabbit fibroma virus, Shope fibroma virus
C:Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 07-May-1999
C:Accession: H41700
R:Strayer, D.S.; Jerng, H.H.; O'Connor, K.
Virology 185, 585-595, 1991
A:Title: Sequence and analysis of a portion of the genomes of Shope fibroma virus and
A:Reference number: A41700; MUID:92074222
A:Accession: H41700
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-144 <STR>
A:Cross-references: GB:M32743

```

```

Query Match      80.0%; Score 32; DB 2; Length 144;
Best Local Similarity 57.1%; Pred. No. 23;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 KKRILHC 7
    ||: |||
Db 73 KKNVHC 79

```

RESULT 12

A38117
hypothetical protein 1 - Anabaena sp. insertion sequence IS895
C:Species: Anabaena sp.
C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 15-Oct-1999
C:Accession: A38117
R:Alam, J.; Vrba, J.M.; Cai, Y.; Martin, J.A.; Weislo, L.J.; Curtis, S.E.
J. Bacteriol. 173, 5778-5783, 1991
A:Title: Characterization of the IS895 family of insertion sequences from the cyanobacterium Anabaena sp.
A:Reference number: A38117; MUID:91358370
A:Accession: A38117
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <ALA>
A:Cross-references: GB:M67475; NID:gl42026; PIDN:AAA98138.1; PID:gl42027
A:Experimental source: strain PCC 7120
C:Genetics:
A:Mobile element: insertion sequence IS895

Query Match 80.0%; Score 32; DB 2; Length 189;
Best Local Similarity 71.4%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|||: ||

Db 171 KKRILHC 177

RESULT 13

AD2414
transposase all4868 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AD2414
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp.
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2414
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB76567.1; PID:gl7134005; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all4868

Query Match 80.0%; Score 32; DB 2; Length 189;
Best Local Similarity 71.4%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|||: ||

Db 171 KKRILHC 177

RESULT 14

AE2064
transposase all2067 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AE2064
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp.
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2064

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:44:48 ; Search time 0.585209 Seconds
(without alignments)
292.168 Million cell updates/sec

Title: us-09-606-129a-34

Perfect score: 40

Sequence: 1 KKRILHC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	85.0	269	4	US-09-006-783A-5
2	34	85.0	294	2	US-09-258-371-10
3	34	85.0	294	3	US-08-751-230-10
4	34	85.0	294	4	US-09-499-082-10
5	34	85.0	294	4	US-05-258-372-10
6	34	85.0	294	4	US-09-006-783A-3
7	33	82.5	527	4	US-09-659-166-2
8	33	82.5	528	2	US-08-802-466-2
9	33	82.5	528	4	US-09-350-484-2
10	32	80.0	565	2	US-08-892-770-6
11	32	80.0	571	2	US-08-892-770-5
12	31	77.5	40	1	US-07-822-566B-4
13	31	77.5	88	1	US-08-336-618-19
14	31	77.5	99	5	PCT-US92-03993-5
15	31	77.5	141	1	US-07-822-966B-6
16	31	77.5	141	2	US-08-803-899-6
17	31	77.5	142	1	US-08-336-618-18
18	31	77.5	142	5	PCT-US92-03993-7
19	31	77.5	257	2	US-08-715-204-6
20	31	77.5	257	3	US-09-162-597-6
21	31	77.5	504	3	US-09-231-529-6
22	31	77.5	504	4	US-08-977-816-6
23	31	77.5	508	2	US-08-818-024-3
24	31	77.5	508	4	US-09-334-775A-3
25	30	75.0	55	1	US-08-469-427A-3
26	30	75.0	55	2	US-08-609-443B-3
27	30	75.0	55	2	US-08-569-063C-3

28	30	75.0	188	1	US-08-469-427A-5	Sequence 5, Appli
29	30	75.0	188	2	US-08-609-443B-5	Sequence 5, Appli
30	30	75.0	188	2	US-08-569-063C-5	Sequence 5, Appli
31	30	75.0	195	1	US-08-469-427A-7	Sequence 7, Appli
32	30	75.0	195	2	US-08-609-443B-7	Sequence 7, Appli
33	30	75.0	195	2	US-08-569-063C-7	Sequence 7, Appli
34	30	75.0	396	3	US-09-082-310-2	Sequence 2, Appli
35	30	75.0	568	2	US-08-835-170-4	Sequence 4, Appli
36	30	75.0	568	4	US-09-359-257-4	Sequence 4, Appli
37	30	75.0	588	2	US-08-835-170-2	Sequence 2, Appli
38	30	75.0	588	4	US-09-359-257-2	Sequence 2, Appli
39	29	72.5	377	2	US-08-969-106-6	Sequence 6, Appli
40	29	72.5	377	4	US-09-054-492B-1	Sequence 1, Appli
41	28	70.0	22	1	US-08-484-635-166	Sequence 166, App
42	28	70.0	22	2	US-08-484-631-166	Sequence 166, App
43	28	70.0	22	2	US-08-827-570-166	Sequence 166, App
44	28	70.0	34	4	US-08-974-549A-200	Sequence 200, App
45	28	70.0	69	3	US-08-851-843A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1

US-09-006-783A-5

; Sequence 5, Application US/09006783A

; Patent No. 6297366

; GENERAL INFORMATION:

; APPLICANT: Gudkov, Andrey V

; APPLICANT: Garkavstev, Igor

; APPLICANT: Riabowol, Karl

; TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling

; TITLE OF INVENTION: Pathway

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff

; STREET: 300 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/006.783A

; FILING DATE: 15-JAN-1998

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 6297366nan, Kevin E

; REGISTRATION NUMBER: 35,303

; REFERENCE/DOCKET NUMBER: 97,837

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-913-0001

; TELEFAX: 312-913-0002

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 269 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-006-783A-5

Query Match 85.0%; Score 34; DB 4; Length 269;

Best Local Similarity 71.4%; Pred. No. 13;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7

Db 67 KRRMLHC 73

RESULT 2
US-09-258-371-10
; Sequence 10, Application US/09258371
; Patent No. 5986078
; GENERAL INFORMATION:
; APPLICANT: Garkavtsev, Igor
; APPLICANT: Riabowol, Karl
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
; TITLE OF INVENTION: SUPPRESSOR GENE INGI
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/258,371
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/751,230
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mool, Leslie A.
; REGISTRATION NUMBER: 37,047
; REFERENCE/DOCKET NUMBER: 028722-144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-7400
; TELEFAX: 415-854-8275
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-258-371-10

Query Match 85.0% Score 34; DB 2; Length 294;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7
Db 82 KRRMLHC 88

RESULT 3
US-08-751-230-10
; Sequence 10, Application US/08751230
; Patent No. 6117633
; GENERAL INFORMATION:
; APPLICANT: Garkavtsev, Igor
; APPLICANT: Riabowol, Karl
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
; TITLE OF INVENTION: SUPPRESSOR GENE INGI
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,230
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/569721
; FILING DATE: 08-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mool, Leslie A.
; REGISTRATION NUMBER: 37,047
; REFERENCE/DOCKET NUMBER: 028722-144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-7400
; TELEFAX: 415-854-8275
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-230-10

Query Match 85.0% Score 34; DB 3; Length 294;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7
Db 82 KRRMLHC 88

RESULT 4
US-09-499-082-10
; Sequence 10, Application US/09499082
; Patent No. 6143522
; GENERAL INFORMATION:
; APPLICANT: Helbing, Caren C.
; APPLICANT: Riabowol, Karl
; APPLICANT: Johnston, Randall N.
; APPLICANT: Garkavtsev, Igor
; TITLE OF INVENTION: METHODS OF MODULATING APOPTOSIS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/499,082
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,158
; FILING DATE: 27-MAR-1997
; APPLICATION NUMBER: US 08/751230
; FILING DATE: 15-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/569721
; FILING DATE: 08-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mool, Leslie A.
; REGISTRATION NUMBER: 37,047
; REFERENCE/DOCKET NUMBER: 028722-148

TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-854-7400
TELEFAX: 650-854-8275
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-499-082-10

Query Match 85.0%; Score 34; DB 4; Length 294;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|:|:|:|
DB 82 KRRMLHC 88

RESULT 5

US-09-258-372-10
Sequence 10, Application US/09258372
Patent No. 6238918
GENERAL INFORMATION:
APPLICANT: Garkavtsev, Igor
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
SUPPRESSOR GENE INGI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/258,372
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/751,230
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mool, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-144
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-258-372-10

Query Match 85.0%; Score 34; DB 4; Length 294;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|:|:|:|
DB 82 KRRMLHC 88

RESULT 6

US-09-006-783A-3
Sequence 3, Application US/09006783A
Patent No. 6297366
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrey V
APPLICANT: Garkavtsev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
TITLE OF INVENTION: Pathway
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,783A
FILING DATE: 15-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6297366nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 97,837
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-006-783A-3

Query Match 85.0%; Score 34; DB 4; Length 294;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|:|:|:|
DB 82 KRRMLHC 88

RESULT 7

US-09-659-166-2
Sequence 2, Application US/09659166
Patent No. 6355465
GENERAL INFORMATION:
APPLICANT: CREASY, CARETHA LEE
APPLICANT: TESTA, TANIA TAMSON
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: Gp-30218
CURRENT APPLICATION NUMBER: US/09/659,166
CURRENT FILING DATE: 2000-09-11
PRIOR APPLICATION NUMBER: UK 9921505.5
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 527
TYPE: PRT
ORGANISM: rattus
US-09-659-166-2

Query Match 82.5%; Score 33; DB 4; Length 527;

Best Local Similarity 71.4%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRILHC 7
| ||| ||
Db 267 KNRIHC 273

RESULT 8
US-08-802-466-2
; Sequence 2, Application US/08802466
; Patent No. 5972606
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,466
; FILING DATE: 19 February 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 528 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-802-466-2

Query Match 82.5%; Score 33; DB 2; Length 528;
Best Local Similarity 71.4%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRILHC 7
| ||| ||
Db 268 KNRIHC 274

RESULT 9
US-09-350-484-2
; Sequence 2, Application US/09350484
; Patent No. 6159716
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia

Query Match 82.5%; Score 33; DB 2; Length 528;
Best Local Similarity 71.4%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRILHC 7
| ||| ||
Db 268 KNRIHC 274

RESULT 10
US-08-892-770-6
; Sequence 6, Application US/08892770
; Patent No. 5962670
; GENERAL INFORMATION:
; APPLICANT: Walling, Linda L.
; APPLICANT: Pautot, Veronique
; APPLICANT: Gu, Yong-Qiang
; APPLICANT: Chao, Wun Shaw
; TITLE OF INVENTION: Improved Promoters for Enhancing Plant
; PRODUCTIVITY
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892,770
; FILING DATE: 15-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261

REFERENCE/DOCKET NUMBER: 023070-072100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..565
OTHER INFORMATION: /note= "deduced partial protein sequence
of acidic leucine aminopeptidase 2
OTHER INFORMATION: (LapA2) from tomato"
US-08-892-770-6

Query Match 80.0%; Score 32; DB 2; Length 565;
Best Local Similarity 83.3%; Pred. No. 63;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRILHC 7
|||:|
Db 42 KRIVHC 47

RESULT 11
US-08-892-770-5
Sequence 5, Application US/08892770
Patent No. 5962670
GENERAL INFORMATION:
APPLICANT: Walling, Linda L.
APPLICANT: Pautot, Veronique
APPLICANT: Gu, Yong-Qiang
APPLICANT: Chao, Wun Shaw
TITLE OF INVENTION: Improved Promoters for Enhancing Plant
PRODUCTIVITY
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/892,770
FILING DATE: 15-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 023070-072100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 571 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..571

OTHER INFORMATION: /note= "deduced full-length protein
sequence of acidic leucine
OTHER INFORMATION: aminopeptidase (LapA1) protein from
tomato"
US-08-892-770-5

Query Match 80.0%; Score 32; DB 2; Length 571;
Best Local Similarity 83.3%; Pred. No. 64;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRILHC 7
|||:|
Db 48 KRIVHC 53

RESULT 12
US-07-822-966B-4
Sequence 4, Application US/07822966B
Patent No. 5498597
GENERAL INFORMATION:
APPLICANT: Steven J. Burakoff
APPLICANT: Stuart L. Schreiber
APPLICANT: Barbara E. Bierer
TITLE OF INVENTION: FKBP-13, AN FK506-BINDING
PROTEIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/822,966B
FILING DATE: January 17, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/052001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 40
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-07-822-966B-4

Query Match 77.5%; Score 31; DB 1; Length 40;
Best Local Similarity 71.4%; Pred. No. 8.4;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRILHC 7
|||:|
Db 14 KKRVDHC 20

RESULT 13
US-08-336-618-19
Sequence 19, Application US/08336618
Patent No. 5763590
GENERAL INFORMATION:
APPLICANT: Peattie, Debra A.
APPLICANT: Harding, Matthew W.

```
; APPLICANT: Livingston, David J.
; TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
; TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
; TITLE OF INVENTION: CDNA
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,618
; FILING DATE: 09-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/963,325
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/777,752
; FILING DATE: 11-OCT-1991
; PRIOR APPLICATION DATA: PCT/
; FILING DATE: 09-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: VP191-06A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-336-618-19

Query Match 77.5%; Score 31; DB 1; Length 88;
Best Local Similarity 71.4%; Pred. No. 18;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRILHC 7
Db 14 KKRVDHC 20

RESULT 14
PCT-US92-03993-5
; Sequence 5, Application PC/TUS9203993
; GENERAL INFORMATION:
; APPLICANT: Harding, Matthew W.
; TITLE OF INVENTION: FKBP: A NOVEL PROLYL ISOMERASE AND
; TITLE OF INVENTION: RAPAMYCIN/FK506 BINDING PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/03993
; FILING DATE: 19920507
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/697,113
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: VP191-05A PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 99 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-03993-5

Query Match 77.5%; Score 31; DB 5; Length 99;
Best Local Similarity 71.4%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRILHC 7
Db 14 KKRVDHC 20

RESULT 15
US-07-822-966B-6
; Sequence 6, Application US/07822966B
; Patent No. 5498597
; GENERAL INFORMATION:
; APPLICANT: Steven J. Burakoff
; APPLICANT: Stuart L. Schreiber
; APPLICANT: Barbara E. Bierer
; TITLE OF INVENTION: FKBP-13, AN FK506-BINDING
; TITLE OF INVENTION: IMMUNOPHILIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/822,966B
; FILING DATE: January 17, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/052001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-822-966B-6
```

Query Match 77.5%; Score 31; DB 1; Length 141;
Best Local Similarity 71.4%; Pred. No. 27;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRILHC 7
| | | | |
Db 35 KKRVDHC 41

Search completed: November 1, 2002, 14:51:58
Job time : 0.585209 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:40:53 ; Search time 1.58682 Seconds
(without alignments)
489.985 Million cell updates/sec

Title: US-09-606-129A-34

Perfect score: 40

Sequence: 1 KKRILHC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	95.0	285	21	Arabidopsis thalia
2	36	90.0	94	21	Murine INGbl isofo
3	36	90.0	279	21	Murine P37INGl pol
4	35	87.5	508	21	Arabidopsis thalia
5	35	87.5	533	21	Arabidopsis thalia
6	35	87.5	564	21	Arabidopsis thalia
7	34	85.0	94	21	Human INGbl isofo
8	34	85.0	128	20	Amino acid sequenc
9	34	85.0	235	22	Amino acid sequenc
10	34	85.0	279	21	Human P37INGl. HO
11	34	85.0	279	22	Tumour suppressor

12	34	85.0	279	22	AAB84698	Amino acid sequenc
13	34	85.0	294	18	AAW19119	Tumour suppressor
14	34	85.0	294	19	AAW79675	Human P33INGl poly
15	34	85.0	294	20	AAV03244	Amino acid sequenc
16	34	85.0	294	22	AAB84697	Amino acid sequenc
17	34	85.0	455	22	ABB71982	Drosophila melanog
18	33	82.5	82	22	AAW87522	Human immune/haema
19	33	82.5	141	21	AAW53247	Human colon cancer
20	33	82.5	350	22	ABW50167	Human transcriptio
21	33	82.5	398	22	ABW62446	Drosophila melanog
22	33	82.5	528	19	AAW64559	Human protein kina
23	33	82.5	528	22	AAE02011	Human YAK1 (hYAK1)
24	33	82.5	549	21	AAW68781	Amino acid sequenc
25	33	82.5	882	22	AAG63551	A human alpha-2 ma
26	33	82.5	899	22	AAG63550	A human alpha-2 ma
27	33	82.5	912	22	AAG63549	A human alpha-2 ma
28	33	82.5	1475	22	ABG21224	Novel human diagno
29	33	82.5	1508	22	AAG63548	A human alpha-2 ma
30	32	80.0	83	22	ABG14241	Novel human diagno
31	32	80.0	126	22	ABG11998	Novel human diagno
32	32	80.0	158	21	AAW25386	Pinus radiata cell
33	32	80.0	565	20	AAW96316	Acidic leucine ami
34	32	80.0	571	20	AAW96315	Acidic leucine ami
35	31	77.5	40	17	AAW93552	Bovine FKBP-13 imm
36	31	77.5	83	22	AAG76114	Human colon cancer
37	31	77.5	93	22	ABG20736	Novel human diagno
38	31	77.5	99	13	AAW28979	Bovine RPKBP. Bos
39	31	77.5	104	21	AAW03758	Human secreted pro
40	31	77.5	122	22	AAW88444	Human immune/haema
41	31	77.5	137	22	AAW87656	Bovine mammary tis
42	31	77.5	141	17	AAW93551	Human FKBP-13 immu
43	31	77.5	182	22	AAU31050	Novel human secret
44	31	77.5	504	18	AAW36140	Bovine P58 protein
45	31	77.5	504	22	ABG09790	Novel human diagno

ALIGNMENTS

RESULT 1

AAW28710
ID AAG28710 standard; Protein: 285 AA.

AC AAG28710;
XX

DT 17-OCT-2000 (first entry)
XX

XX Arabidopsis thaliana protein fragment SEQ ID NO: 34031.
DE

XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

OS Arabidopsis thaliana.
XX

PN EPI033405-A2.
XX

PD 06-SEP-2000.
XX

PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
PR

PR 05-MAR-1999; 99US-0123180.
PR

PR 09-MAR-1999; 99US-0123548.
PR

PR 23-MAR-1999; 99US-0125788.
PR

PR 25-MAR-1999; 99US-0126264.
PR

PR 29-MAR-1999; 99US-0126785.
PR

PR 01-APR-1999; 99US-0127462.
PR

PR 06-APR-1999; 99US-0128234.
PR

PR 08-APR-1999; 99US-0128714.
PR

PR 16-APR-1999; 99US-0129845.
PR

PR 19-APR-1999; 99US-0130077.
PR

PR 21-APR-1999; 99US-0130449.
PR

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132046.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 22-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 28-OCT-1999; 99US-0162142.

Query Match 95.0%; Score 38; DB 21; Length 285;

Best Local Similarity 85.7%; Pred. No. 10;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7
 Db 13 KKRLLHC 19

RESULT 2

ID AAY97243 standard; Protein; 94 AA.

AC AAY97243;

DT 19-DEC-2000 (first entry)

DE Murine ING1 isoform N-terminal peptide.

XX p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
 KW ING1; ing1; p33ING1; p37ING1; oncogene; gene therapy; diagnosis;
 KW proliferation disorder; transformation; transformed cell; mouse.
 XX Mus musculus.

OS Mus musculus.

XX WO200046370-A1.

XX 10-AUG-2000.

XX 04-FEB-2000; 2000WO-US02959.

XX 04-FEB-1999; 99US-0118941.

XX (UNII) UNIV ILLINOIS FOUND.

XX Gudkov A, Zeremski M, Gurova KV, Grigorian IA;

XX WPI; 2000-491278/43.

XX Detecting nucleic acid encoding exon 1b of ing1, useful for diagnosing
 PT and treating cancer, comprises contacting sample with isolated nucleic
 PT acid comprising sequence of exon 1b and detecting hybridized products

XX Claim 6; Fig 7a; 134pp; English.

XX Mutations in or loss of the p53 gene occur in more than 50% of
 CC human tumours and tumour cell lines, but functional inactivation of
 CC the p53 pathway occurs in a much larger proportion of tumours. In
 CC many cases the mechanism of functional inactivation of the p53 gene
 CC remains unknown but p53 has been found to act in cooperation with

CC ING1. Functional cooperation between ING1 and p53 suggested that
 CC ING1 encoded a tumour suppressor protein that functioned within the
 CC p53 pathway. This data suggested a possible role for ING1 in head
 CC and neck cancers and chromosomal location of the ING1 placed it
 CC within a region that is frequently rearranged in head and neck
 CC cancers. Large scale analysis of tumours involving ING1 has not
 CC revealed mutations in ING1 nor significant variations in its
 CC expression suggesting that ING1 was not a useful gene to study in
 CC cancer etiology. However, alternative initiation exons of the ING1
 CC gene, each having their own promoter have been discovered.
 CC Expression of one promoter (1a) produces a protein identical to
 CC ING1. Expression of a second promoter (1b) produces a protein having
 CC an identical C-terminal fragment to ING1 but an additional 104
 CC N-terminal amino acids. The newly discovered protein has been
 CC designated p37ING1 (Wild type: p33ING1). p37ING1 has the
 CC characteristics of an oncogene. When overexpressed in cells (even
 CC those expressing wild type p53) p37ING1 is able to cause
 CC proliferation or transformation of those cells. Thus detecting a
 CC nucleic acid encoding exon 1b of ING1 by hybridisation with an
 CC isolated nucleic acid having the sequence of exon 1b of ING1
 CC or its antisense sequence can identify individuals expressing the
 CC oncogenic form of ING1. Novel peptide sequences taken from the 104
 CC N-terminal peptide of p37ING1 can also be used to raise antibodies
 CC that can also be used in detection methods for the p37ING1 variant.
 CC The polypeptides may be useful in gene therapy for treatment of cell
 CC proliferation disorders, especially cancers and for diagnosing and
 CC studying cancers.

XX Sequence 94 AA;

Query Match 90.0%; Score 36; DB 21; Length 94;

Best Local Similarity 71.4%; Pred. No. 8.8;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7

Db 67 KRRVLHC 73

RESULT 3

ID AAY97242 standard; Protein; 279 AA.

XX AAY97242;

XX 19-DEC-2000 (first entry)

XX Murine p37ING1 polypeptide.

XX p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
 KW ING1; ing1; p33ING1; p37ING1; oncogene; gene therapy; diagnosis;
 KW proliferation disorder; transformation; transformed cell; mouse.

XX Mus musculus.

XX WO200046370-A1.

XX 10-AUG-2000.

XX 04-FEB-2000; 2000WO-US02959.

XX 04-FEB-1999; 99US-0118941.

XX (UNII) UNIV ILLINOIS FOUND.

XX Gudkov A, Zeremski M, Gurova KV, Grigorian IA;

XX WPI; 2000-491278/43.

XX N-PSDB; AAA53790.

XX Detecting nucleic acid encoding exon 1b of ing1, useful for diagnosing
 PT and treating cancer, comprises contacting sample with isolated nucleic
 PT acid comprising sequence of exon 1b and detecting hybridized products

XX Disclosure; Fig 12; 134pp; English.

XX Mutations in or loss of the p53 gene occur in more than 50% of

CC human tumours and tumour cell lines, but functional inactivation of

CC the p53 pathway occurs in a much larger proportion of tumours. In

CC many cases the mechanism of functional inactivation of the p53 gene

CC remains unknown but p53 has been found to act in cooperation with

CC INK1. Functional cooperation between INK1 and p53 suggested that

CC INK1 encoded a tumour suppressor protein that functioned within the

CC p53 pathway. This data suggested a possible role for INK1 in head

CC and neck cancers and chromosomal location of the INK1 placed it

CC within a region that is frequently rearranged in head and neck

CC cancers. Large scale analysis of tumours involving INK1 has not

CC revealed mutations in INK1 nor significant variations in its

CC expression suggesting that INK1 was not a useful gene to study in

CC cancer etiology. However, alternative initiation exons of the INK1

CC gene, each having their own promoter have been discovered.

CC Expression of one promoter (1a) produces a protein identical to

CC INK1. Expression of a second promoter (1b) produces a protein having

CC an identical C-terminal fragment to INK1 but an additional 104

CC N-terminal amino acids. The newly discovered protein has been

CC designated p37ING1 (Wild type: p33ING1). p37ING1 has the

CC characteristics of an oncogene. When overexpressed in cells (even

CC those expressing wild type p53) p37ING1 is able to cause

CC proliferation or transformation of those cells. Thus detecting a

CC nucleic acid encoding exon 1b of INK1 by hybridisation with an

CC isolated nucleic acid having the sequence of exon 1b of INK1

CC or its antisense sequence can identify individuals expressing the

CC oncogenic form of INK1. Novel peptide sequences taken from the 104

CC N-terminal peptide of p37ING1 can also be used to raise antibodies

CC that can also be used in detection methods for the p37ING1 variant.

CC The polypeptides may be useful in gene therapy for treatment of cell

CC proliferation disorders, especially cancers and for diagnosing and

CC studying cancers.

XX SQ Sequence 279 AA;

Query Match 90.0%; Score 36; DB 21; Length 279;

Best Local Similarity 71.4%; Pred. No. 24;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7

Db 67 KRRVLHC 73

RESULT 4

AAG48146

XX AAG48146 standard; Protein: 508 AA.

XX AC AAG48146;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60770.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123348.

XX PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151920.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 87.5% Score 35; DB 21; Length 508;

Best Local Similarity 71.4% Pred. No. 67;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7
|:|:|
Db 236 KRRLHC 242

RESULT 5

AAG48145
ID AAG48145 standard; Protein; 533 AA.

XX AC AAG48145;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60769.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999;	99US-0130449.	PR 20-JUL-1999;	99US-0144352.
PR 23-APR-1999;	99US-0130510.	PR 20-JUL-1999;	99US-0144632.
PR 23-APR-1999;	99US-0130891.	PR 20-JUL-1999;	99US-0144884.
PR 28-APR-1999;	99US-0131449.	PR 21-JUL-1999;	99US-0144814.
PR 30-APR-1999;	99US-0132048.	PR 21-JUL-1999;	99US-0145086.
PR 30-APR-1999;	99US-0132407.	PR 21-JUL-1999;	99US-0145088.
PR 04-MAY-1999;	99US-0132484.	PR 22-JUL-1999;	99US-0145085.
PR 05-MAY-1999;	99US-0132485.	PR 22-JUL-1999;	99US-0145087.
PR 06-MAY-1999;	99US-0132486.	PR 22-JUL-1999;	99US-0145089.
PR 06-MAY-1999;	99US-0132487.	PR 22-JUL-1999;	99US-0145089.
PR 07-MAY-1999;	99US-0132863.	PR 23-JUL-1999;	99US-0145192.
PR 11-MAY-1999;	99US-0134256.	PR 23-JUL-1999;	99US-0145218.
PR 14-MAY-1999;	99US-0134218.	PR 23-JUL-1999;	99US-0145224.
PR 14-MAY-1999;	99US-0134219.	PR 26-JUL-1999;	99US-0145276.
PR 14-MAY-1999;	99US-0134221.	PR 27-JUL-1999;	99US-0145913.
PR 14-MAY-1999;	99US-0134370.	PR 27-JUL-1999;	99US-0145918.
PR 18-MAY-1999;	99US-0134768.	PR 27-JUL-1999;	99US-0145919.
PR 19-MAY-1999;	99US-0134941.	PR 28-JUL-1999;	99US-0145951.
PR 20-MAY-1999;	99US-0135124.	PR 02-AUG-1999;	99US-0146386.
PR 21-MAY-1999;	99US-0135353.	PR 02-AUG-1999;	99US-0146388.
PR 24-MAY-1999;	99US-0135629.	PR 02-AUG-1999;	99US-0146389.
PR 25-MAY-1999;	99US-0136021.	PR 03-AUG-1999;	99US-0147038.
PR 27-MAY-1999;	99US-0136352.	PR 04-AUG-1999;	99US-0147204.
PR 28-MAY-1999;	99US-0136782.	PR 04-AUG-1999;	99US-0147302.
PR 01-JUN-1999;	99US-0137222.	PR 05-AUG-1999;	99US-0147192.
PR 03-JUN-1999;	99US-0137528.	PR 05-AUG-1999;	99US-0147260.
PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 11-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151065.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 30-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140895.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0156659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0158293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.

```
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      87.5%  Score 35;  DB 21;  Length 533;
Best Local Similarity 71.4%  Pred. No. 70;
Matches 5;  Conservative 2;  Mismatches 0;  Indels 0;  Gaps 0;

OY 1 KRRLHC 7
    I: I I I I
Db 261 KRRLHC 267

RESULT 6
AAG48144
ID AAG48144 standard; Protein; 564 AA.
XX
AC AAG48144;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60768.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 05-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 25-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136592.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
```

PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.

PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 87.5%; Score 35; DB 21; Length 564;
Best Local Similarity 71.4%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRLLHC 7

|:|:|

Db 292 KKRLLHC 298

RESULT 7

AA97245

ID AAY97245 standard; Protein; 94 AA.

XX AAY97245;

AC AAY97245;

XX AAY97245;

DT 19-DEC-2000 (first entry)

XX Human INGB1 isoform N-terminal peptide.

XX p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
KW INGL1; ingl1; p31ING1; p37ING1; oncogene; gene therapy; diagnosis;
KW proliferation disorder; transformation; transformed cell; mouse.
XX Homo sapiens.

OS WO200046370-A1.

XX 10-AUG-2000.

XX 04-FEB-2000; 2000WO-US02959.

XX 04-FEB-1999; 99US-0118941.

XX (UNII) UNIV ILLINOIS FOUND.

XX Gudkov A, Zeremski M, Gurova KV, Grigorian IA;

XX WPI; 2000-491278/43.

XX Detecting nucleic acid encoding exon 1b of ingl1, useful for diagnosing

XX and treating cancer, comprises contacting sample with isolated nucleic

XX acid comprising sequence of exon 1b and detecting hybridized products

XX Clalm 26; Figure 7a; 134pp; English.

XX Mutations in or loss of the p53 gene occur in more than 50% of

XX human tumours and tumour cell lines, but functional inactivation of

XX the p53 pathway occurs in a much larger proportion of tumours. In

XX many cases the mechanism of functional inactivation of the p53 gene

XX remains unknown but p53 has been found to act in cooperation with

XX INGL1. Functional cooperation between INGL1 and p53 suggested that

XX INGL1 encoded a tumour suppressor protein that functioned within the

XX p53 pathway. This data suggested a possible role for INGL1 in head

XX and neck cancers and chromosomal location of the INGL1 placed it

XX within a region that is frequently rearranged in head and neck

XX cancers. Large scale analysis of tumours involving INGL1 has not

CC revealed mutations in INGI nor significant variations in its
 CC expression suggesting that INGI was not a useful gene to study in
 CC cancer etiology. However, alternative initiation exons of the ingl
 CC gene, each having their own promoter have been discovered.
 CC Expression of one promoter (la) produces a protein identical to
 CC INGI. Expression of a second promoter (lb) produces a protein having
 CC an identical C-terminal fragment to INGI but an additional 104
 CC N-terminal amino acids. The newly discovered protein has been
 CC designated p37ING1 (Wild type: p33ING1). p37ING1 has the
 CC characteristics of an oncogene. When overexpressed in cells (even
 CC those expressing wild type p53) p37ING1 is able to cause
 CC proliferation or transformation of those cells. Thus detecting a
 CC nucleic acid encoding exon lb of ingl by hybridisation with an
 CC isolated nucleic acid having the sequence of exon lb of ingl
 CC or its antisense sequence can identify individuals expressing the
 CC oncogenic form of ingl. Novel peptide sequences taken from the 104
 CC N-terminal peptide of p37ING1 can also be used to raise antibodies
 CC that can also be used in detection methods for the p37ING1 variant.
 CC The polypeptides may be useful in gene therapy for treatment of cell
 CC proliferation disorders, especially cancers and for diagnosing and
 CC studying cancers.

SQ Sequence 94 AA;

Query Match 85.0%; Score 34; DB 21; Length 94;
 Best Local Similarity 71.4%; Pred. No. 22;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
 |:|:|:|
 Db 67 KRRMLHC 73

RESULT 8
 AAY03245

ID AAY03245 standard; Protein: 128 AA.

XX AAY03245;

XX 26-AUG-1999 (first entry)

DE Amino acid sequence of the human p28-ING1 protein.

KW Human; p33-ING1 protein; growth regulation; apoptosis; DNA damage;
 KW inhibition; anchorage independent growth; cytotoxic drug; cancer;
 KW transcriptional activation; immortal cell line; p28-ING1 protein.

XX Homo sapiens.

XX Key Location/Qualifiers

FT FT Difference 59..269

FT Misc_Difference /note= "p26-ING1 fragment"

XX WO9916790-A1.

XX 08-APR-1999.

XX 24-SEP-1998; 98WO-US18179.

XX 14-JAN-1998; 98US-0006783.

PR 26-SEP-1997; 97US-0060138.

XX (UNII) UNIV ILLINOIS BOARD OF TRUSTEES.

PA (UYTE-) UNIV TECHNOLOGIES INT INC.

XX Garkavtsev I, Gudkov A, Riabowol K;

XX WPI; 1999-263685/22.

XX Use of p33-ING1 peptides

XX Example 8; Page 61; 64pp; English.

CC This is the amino acid sequence of the human p28-ING1 protein,
 CC used in the method of the invention, involving the human p33-ING1
 CC protein. The INGI gene encodes p33-ING1 which can be used to
 CC modulate the activity of, isolate or detect p53. Expression of the
 CC INGI and p53 genes in a mammalian cell results in normal growth
 CC regulation anchorage-dependent growth and apoptosis as a response
 CC to irreversible DNA damage and other cellular insult. Inhibition of
 CC expression of either gene results in a loss of cellular growth
 CC control, anchorage independent growth, inhibition of apoptosis
 CC and resistance to radiation and cytotoxic drugs. The p33-ING1 is a
 CC component of the p53 signalling pathway that cooperates with p53 in
 CC negative regulation of cell proliferation by modulating p53 dependent
 CC transcriptional activation. Biological function of p53 signalling
 CC pathway can therefore be regulated (both enhanced or suppressed) by
 CC modulating p33-ING1 activity. The modulation of p33-ING1 activity can
 CC be used for the stimulation or restoration of the p53 pathway in
 CC anti cancer therapy or for the suppression of the p53 pathway to
 CC defend sensitive tissues from genotoxic stress or for the generation
 CC of immortal cell lines.

XX Sequence 128 AA;

Query Match 85.0%; Score 34; DB 20; Length 128;

Best Local Similarity 71.4%; Pred. No. 29;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|:|:|:|
 Db 67 KRRMLHC 73

RESULT 9

AAB84700

ID AAB84700 standard; Protein: 235 AA.

XX AAB84700;

XX 17-SEP-2001 (first entry)

DE Amino acid sequence of a human cancer associated antigen.

XX Cancer associated antigen; INGI; tumour suppressor; cancer; vaccine.

XX Homo sapiens.

XX WO200147959-A2.

XX 05-JUL-2001.

XX 29-NOV-2000; 2000WO-US42334.

XX 30-NOV-1999; 99US-0451739.

PR 24-OCT-2000; 2000US-0602362.

XX (LUDW-) LUDWIG INST CANCER RES.

PA (SLOK) SLOAN KETTERING INST CANCER RES.

PA (CORR) CORNELL RES FOUND INC.

XX Jager D, Stockert E, Scanlan M, Knuth A, Old L, Gure A, Chen Y;

XX WPI; 2001-441706/47.

XX Isolated cancer associated nucleic acid molecule identified by SEREX
 PT (serological identification of antigens by recombinant expression
 PT cloning) technique, useful in nucleic acid based therapies to treat
 PT cancer -

XX Claim 83; Page 47-48; 62pp; English.

XX The present sequence represents a human cancer associated antigen.

CC The sequence is a variant of the INGI gene, which is a tumour

CC suppressor gene candidate. The cancer associated antigen polynucleotides

CC and polypeptides are useful for screening for the possible presence of

CC a pathological condition in a subject such as cancer. The cancer
 CC associated antigen polypeptides are useful for producing vaccines.
 XX
 SQ Sequence 235 AA;

Query Match 85.0%; Score 34; DB 22; Length 235;
 Best Local Similarity 71.4%; Pred. No. 51;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
 |:|:|:|
 Db 23 KRRMLHC 29

RESULT 10
 AAY97244
 ID AAY97244 standard; Protein; 279 AA.

XX AAY97244;
 XX 19-DEC-2000 (first entry)
 DT
 XX Human P37ING1.

XX p53; tumour; cancer; detection; antibody; hybridisation; exon 1b;
 KW ING1; ing1; p37ING1; p37ING1; oncogene; gene therapy; diagnosis;
 KW proliferation disorder; transformation; transformed cell; mouse.
 XX
 OS Homo sapiens.

XX WO200046370-A1.
 XX 10-AUG-2000.
 XX 04-FEB-2000; 2000WO-US02959.
 XX 04-FEB-1999; 99US-0118941.
 XX (UNII) UNIV ILLINOIS FOUND.
 XX Gudkov A, Zeremski M, Gurova KV, Grigorian IA;
 PI WPI; 2000-491278/43.
 DR N-PSDB; AAA53792.

XX Detecting nucleic acid encoding exon 1b of ing1, useful for diagnosing
 PT and treating cancer, comprises contacting sample with isolated nucleic
 PT acid comprising sequence of exon 1b and detecting hybridized products
 XX
 PS Disclosure; Fig 7a; 134pp; English.

XX Mutations in or loss of the p53 gene occur in more than 50% of
 CC human tumours and tumour cell lines, but functional inactivation of
 CC the p53 pathway occurs in a much larger proportion of tumours. In
 CC many cases the mechanism of functional inactivation of the p53 gene
 CC remains unknown but p53 has been found to act in cooperation with
 CC ING1. Functional cooperation between ING1 and p53 suggested that
 CC ING1 encoded a tumour suppressor protein that functioned within the
 CC p53 pathway. This data suggested a possible role for ING1 in head
 CC and neck cancers and chromosomal location of the ING1 placed it
 CC within a region that is frequently rearranged in head and neck
 CC cancers. Large scale analysis of tumours involving ING1 has not
 CC revealed mutations in ING1 nor significant variations in its
 CC expression suggesting that ING1 was not a useful gene to study in
 CC cancer etiology. However, alternative initiation exons of the ing1
 CC gene, each having their own promoter have been discovered.
 CC Expression of one promoter (1a) produces a protein identical to
 CC ING1. Expression of a second promoter (1b) produces a protein having
 CC an identical C-terminal fragment to ING1 but an additional 104
 CC N-terminal amino acids. The newly discovered protein has been
 CC designated p37ING1 (wild type: p33ING1). p37ING1 has the
 CC characteristics of an oncogene. When overexpressed in cells (even
 CC those expressing wild type p53) p37ING1 is able to cause

CC proliferation or transformation of those cells. Thus detecting a
 CC nucleic acid encoding exon 1b of ing1 by hybridisation with an
 CC isolated nucleic acid having the sequence of exon 1b of ing1
 CC or its antisense sequence can identify individuals expressing the
 CC oncogenic form of ing1. Novel peptide sequences taken from the 104
 CC N-terminal peptide of p37ING1 can also be used to raise antibodies
 CC that can also be used in detection methods for the p37ING1 variant.
 CC The polypeptides may be useful in gene therapy for treatment of cell
 CC proliferation disorders, especially cancers and for diagnosing and
 CC studying cancers.

XX Sequence 279 AA;

Query Match 85.0%; Score 34; DB 21; Length 279;
 Best Local Similarity 71.4%; Pred. No. 60;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
 |:|:|:|
 Db 67 KRRMLHC 73

RESULT 11
 AAE06676
 ID AAE06676 standard; Protein; 279 AA.

XX AAE06676;
 XX 23-OCT-2001 (first entry)
 DT
 XX Tumour suppressor homologue protein, p33ING1.

XX Tumour suppressor; p47ING3; cell proliferation; cellular aging;
 KW anchorage dependence; apoptosis; tumour; cancer; gene therapy; p33ING1.
 KW
 XX Unidentified.

XX WO200159114-A2.
 XX 16-AUG-2001.
 XX 09-FEB-2001; 2001WO-US04425.
 XX 09-FEB-2000; 2000US-0181292.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Harris CC, Nagashima M;
 PI WPI; 2001-488975/53.
 DR
 XX New tumor suppressor protein p47ING3 for the diagnosis and treatment of
 PT tumors -

XX Example 1; Page 78-79; 80pp; English.
 PS The present sequence is a tumour suppressor homologue protein, p33ING1
 CC which is homologous to human tumour suppressor protein, p47ING3. The
 CC tumour suppressors of the invention are involved in the regulation of
 CC cell proliferation and in the control of cellular aging, anchorage
 CC dependence and apoptosis. The tumour suppressor protein, p47ING3,
 CC nucleic acids encoding it and antibodies against it are useful for
 CC diagnosis, prevention and treatment of tumours and cancers. The p47ING3
 CC DNA is also used in gene therapy.

XX Sequence 279 AA;

Query Match 85.0%; Score 34; DB 22; Length 279;
 Best Local Similarity 71.4%; Pred. No. 60;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7
 |:|:|:|

Db 67 KRRMLHC 73

RESULT 12

AA884698
ID AAB84698 standard; Protein: 279 AA.

XX AC AAB84698;

XX DT 17-SEP-2001 (first entry)

XX DE Amino acid sequence of a human cancer associated antigen.

XX KW Cancer associated antigen; INGI1; tumour suppressor; cancer; vaccine.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 123 /note= "Val encoded by GCG"

FT Misc-difference 129 /note= "Val encoded by GCT"

FT Misc-difference 135 /note= "Asn encoded by AAA"

FT Misc-difference 137 /note= "Asp encoded by GAG"

FT Misc-difference 139 /note= "Val encoded by GCA"

FT Misc-difference 142 /note= "Ser encoded by GCT"

XX PN WO200147959-A2.

XX PD 05-JUL-2001.

XX PF 29-NOV-2000; 2000WO-US42334.

XX PR 30-NOV-1999; 99US-0451739.

XX PR 24-OCT-2000; 2000US-0602362.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PA (SLOK) SLOAN KETTERING INST CANCER RES.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Jager D, Stockert E, Scanlan M, Knuth A, Old L, Gure A, Chen Y;

XX DR WPI; 2001-441706/47.

XX DR N-PSDB; AAH28478.

XX PT Isolated cancer associated nucleic acid molecule identified by SEREX

XX PT (serological identification of antigens by recombinant expression

XX PT cloning) technique, useful in nucleic acid based therapies to treat

XX PT cancer -

XX PS Claim 83; Page 45-46; 62pp; English.

XX CC The present sequence represents a human cancer associated antigen.

XX CC The sequence is a variant of the INGI1 gene, which is a tumour

XX CC suppressor gene candidate. The cancer associated antigen polynucleotides

XX CC and polypeptides are useful for screening for the possible presence of

XX CC a pathological condition in a subject such as cancer. The cancer

XX CC associated antigen polypeptides are useful for producing vaccines.

XX SQ Sequence 279 AA;

Query Match

Best Local Similarity 85.0%; Score 34; DB 22; Length 279;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7

Db 67 KRRMLHC 73

RESULT 13

AAW19119

ID AAW19119 standard; Protein: 294 AA.

XX AC AAW19119;

XX DT 27-AUG-1997 (first entry)

XX DE Tumour suppressor protein p33ING1.

XX KW Tumour suppressor protein; p33ING1; breast cancer; brain cancer;

XX KW diagnosis; therapy.

XX OS Homo sapiens.

XX PN WO9721809-A1.

XX PD 19-JUN-1997.

XX PF 06-DEC-1996; 96WO-CA00819.

XX PR 15-NOV-1996; 96US-0751230.

XX PR 08-DEC-1995; 95US-0569721.

XX PA (UYTE-) UNIV TECHNOLOGIES INT INC.

XX PI Garkavstev I, Riabowol K;

XX DR WPI; 1997-332781/30.

XX DR N-PSDB; AAT69652.

XX PT Isolated tumour suppressor gene, INGI1 - useful to develop products

XX PT for inhibiting or increasing cell proliferation, in particular for

XX PT treatment or diagnosis of cancer

XX PS Claim 19; Fig 3; 63pp; English.

XX CC A full-length sequence (AAW19119) is provided of novel tumour

XX CC suppressor protein p33ING1. It was deduced from a cDNA clone

XX CC of p33ING1 (formerly p33IG1) is thought to inhibit cell growth and

XX CC DNA synthesis, probably by arresting cells in the G1 phase of the

XX CC cell cycle. p33ING1 is expressed in normal mammary epithelial

XX CC cells, but expressed only at lower levels in several cancerous

XX CC mammary epithelial cell lines and is not expressed in many primary

XX CC brain tumours. p33ING1 can be administered to patients as a means

XX CC of decreasing proliferation of cancer cells, or can be used to

XX CC raise antibodies useful for the diagnosis of cancer, esp. breast

XX CC and brain cancer.

XX SQ Sequence 294 AA;

Query Match

Best Local Similarity 85.0%; Score 34; DB 18; Length 294;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRILHC 7

Db 82 KRRMLHC 88

RESULT 14

AAW79675

ID AAW79675 standard; Protein: 294 AA.

XX AC AAW79675;

XX DT 18-JAN-1999 (first entry)

XX DE Human p33ING1 polypeptide.

XX KW INGI1 gene; p33ING1; human; apoptosis; cell death; breast cancer;

KW brain tumour; gene therapy; tumour suppressor.

XX Homo sapiens.

XX WO9844102-A2.

XX 08-OCT-1998.

XX 26-MAR-1998; 98WO-CA00277.

XX 27-MAR-1997; 97US-0828158.

XX (UYTE-) UNIV TECHNOLOGIES INT INC.

XX Garkavtsev I, Helbing CC, Johnston RN, Riabowol K;

XX WPI; 1998-542700/46.

XX N-PSDB; AAV62292.

XX Modulating eukaryotic apoptosis by increasing p33ING1 activity -
PT using p33ING1 derivatives, to induce apoptosis in cancer cells, and
PT in the investigation of apoptotic pathways

XX Example 2; Fig 3; 66pp; English.

CC This is the amino acid sequence of human p33ING1, a novel inhibitor
CC of cell growth and a candidate tumour suppressor. The sequence was
CC deduced from full-length ING1 cDNA (see AAV62285) obtained using a
CC positive selection procedure that combined subtractive hybridisation
CC with an in vivo selection assay to identify growth-suppressor
CC elements. ING1 is expressed in normal mammary epithelial cells, but
CC is expressed only at lower levels in several cancerous mammary
CC epithelial cell lines and is not expressed in many primary brain
CC tumours. A claimed method to potentiate apoptosis in a eukaryotic
CC cell involves administering an active p33ING1 peptide or an
CC oligonucleotide encoding such as a peptide. A claimed method for
CC inhibiting apoptosis in a eukaryotic cell involves administering an
CC antisense oligonucleotide. Also claimed are a method for
CC determining the apoptotic characteristics of a eukaryotic cell, an
CC assay for determining the level of p33ING1 activity in a eukaryotic
CC cell, and an isolated eukaryotic cell substantially free of p33ING1
CC biological activity. The invention discloses ING1 derivatives or
CC variants that may be used to induce apoptosis in eukaryotic cancer
CC cells.

XX Sequence 294 AA;

Query Match 85.0%; Score 34; DB 19; Length 294;

Best Local Similarity 71.4%; Pred. No. 63;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|.:|:|

Db 82 KRRMLHC 88

RESULT 15

AAV03244

ID AAV03244 standard; Protein; 294 AA.

XX AAV03244;

XX 26-AUG-1999 (first entry)

XX Amino acid sequence of the human p33-ING1 protein.

XX Human; p33-ING1 protein; growth regulation; apoptosis; DNA damage;

KW inhibition; anchorage independent growth; cytotoxic drug;

KW transcriptional activation; cancer; immortal cell line.

XX Homo sapiens.

OS

XX WO9916790-A1.

PN

XX 08-APR-1999.

XX 24-SEP-1998; 98WO-US18179.

XX 14-JAN-1998; 98US-0006783.

XX 26-SEP-1997; 97US-0060138.

XX (UNII) UNIV ILLINOIS BOARD OF TRUSTEES.

XX (UYTE-) UNIV TECHNOLOGIES INT INC.

XX Garkavtsev I, Gudkov A, Riabowol K;

XX WPI; 1999-263685/22.

XX N-PSDB; AAX28688.

XX Use of p33-ING1 peptides

XX Disclosure; Fig 1; 64pp; English.

XX This is the amino acid sequence of the human p33-ING1 protein.
XX used in the method of the invention. The ING1 gene encodes p33-ING1
XX which can be used to modulate the activity of, isolate or detect p53.
CC Expression of the ING1 and p53 genes in a mammalian cell results in
CC normal growth regulation anchorage-dependent growth and apoptosis as
CC a response to irreversible DNA damage and other cellular insult.
CC Inhibition of expression of either gene results in a loss of cellular
CC growth control, anchorage independent growth, inhibition of apoptosis
CC and resistance to radiation and cytotoxic drugs. The p33-ING1 is a
CC component of the p53 signalling pathway that cooperates with p53 in
CC negative regulation of cell proliferation by modulating p53 dependent
CC transcriptional activation. Biological function of p53 signalling
CC pathway can therefore be regulated (both enhanced or suppressed) by
CC modulating p33-ING1 activity. The modulation of p33-ING1 activity can
CC be used for the stimulation or restoration of the p53 pathway in
CC anti cancer therapy or for the suppression of the p53 pathway to
CC defend sensitive tissues from genotoxic stress or for the generation
CC of immortal cell lines.

XX Sequence 294 AA;

Query Match 85.0%; Score 34; DB 20; Length 294;

Best Local Similarity 71.4%; Pred. No. 63;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRILHC 7

|.:|:|

Db 82 KRRMLHC 88

Search completed: November 1, 2002, 14:47:09

Job time : 2.58682 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:08 ; Search time 1.45338 Seconds
(without alignments)
952.238 Million cell updates/sec

Title: US-09-606-129a-35

Perfect score: 49

Sequence: 1 QKYCCSRK 8

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	49	100.0	296	4 Q9BRW8	Q9brw8 homo sapien
2	49	100.0	296	4 Q96QL4	Q96ql4 homo sapien
3	38	77.6	703	16 Q9KQ22	Q9kq22 vibrio chol
4	38	77.6	982	12 Q9JGP6	Q9jgp6 epizootic h
5	37	75.5	234	4 Q9UIS5	Q9uis5 homo sapien
6	37	75.5	243	4 Q9UIS4	Q9uis4 homo sapien
7	37	75.5	571	5 Q18233	Q18233 caenorhabdi
8	36	73.5	58	5 Q9V8F4	Q9v8f4 drosophila
9	36	73.5	71	8 Q98S56	Q98s56 guillardia
10	36	73.5	150	4 Q15371	Q15371 homo sapien
11	36	73.5	263	5 Q44161	Q44161 caenorhabdi
12	35	71.4	88	10 Q40142	Q40142 lycopersico
13	35	71.4	127	5 Q17778	Q17778 caenorhabdi
14	35	71.4	216	5 Q9BKL4	Q9bkl4 plasmodium
15	35	71.4	330	4 Q9UGG9	Q9ugg9 homo sapien
16	35	71.4	385	5 Q45193	Q45193 caenorhabdi

17	35	71.4	438	10 Q9XEP7	Q9xep7 sorghum bic
18	35	71.4	468	12 Q55279	Q55279 avian adeno
19	35	71.4	959	5 Q9U3D7	Q9u3d7 caenorhabdi
20	35	71.4	961	5 Q9U3D6	Q9u3d6 caenorhabdi
21	35	71.4	3190	5 Q01368	Q01368 drosophila
22	35	71.4	3275	5 Q9W321	Q9w321 drosophila
23	34	69.4	129	3 Q13540	Q13540 saccharomyc
24	34	69.4	158	11 Q9D1Y9	Q9d1y9 mus musculu
25	34	69.4	217	12 Q81864	Q81864 hepatitis e
26	34	69.4	236	16 Q97IK7	Q97ik7 clostridium
27	34	69.4	307	13 Q90ZT5	Q90zt5 xenopus lae
28	34	69.4	485	10 Q9LN19	Q9ln19 arabidopsis
29	34	69.4	491	12 Q9YWL2	Q9ywl2 hepatitis e
30	34	69.4	1484	5 Q94298	Q94298 caenorhabdi
31	34	69.4	1685	12 Q9E8G6	Q9e8g6 hepatitis e
32	34	69.4	1693	12 Q89444	Q89444 hepatitis e
33	34	69.4	1693	12 Q81862	Q81862 hepatitis e
34	34	69.4	1693	12 Q81344	Q81344 hepatitis e
35	34	69.4	1693	12 Q69418	Q69418 hepatitis e
36	34	69.4	1693	12 Q69410	Q69410 hepatitis e
37	34	69.4	1693	12 Q39221	Q39221 hepatitis e
38	34	69.4	1693	12 Q9WLL5	Q9wll5 hepatitis e
39	34	69.4	1693	12 Q9WC28	Q9wc28 hepatitis e
40	34	69.4	1693	12 Q81876	Q81876 hepatitis e
41	34	69.4	1698	12 Q9YLR3	Q9ylr3 hepatitis e
42	34	69.4	1703	12 Q911I6	Q911i6 hepatitis e
43	34	69.4	1707	12 Q91VZ9	Q91vz9 hepatitis e
44	34	69.4	1708	12 Q9YLR1	Q9ylr1 hepatitis e
45	34	69.4	1708	12 Q9YK10	Q9yk10 swine hepat

ALIGNMENTS

RESULT 1

Q9BRW8 ID Q9BRW8 PRELIMINARY; PRT; 296 AA.
AC Q9BRW8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SIMILAR TO BILIVERDIN REDUCTASE A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN, PRIMITIVE NEUROECTODERMAL;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005902; AAH05902.1; -;
DR InterPro; IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
SQ SEQUENCE 296 AA; 33432 MW; A389AFDDB6ADBABF CRC64;

Query Match 100.0%; Score 49; DB 4; Length 296;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKYCCSRK 8
|||||||
Db 289 QKYCCSRK 296

RESULT 2

Q96QL4 ID Q96QL4 PRELIMINARY; PRT; 296 AA.
AC Q96QL4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:14706).

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=PROSTATE;
RC Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC008456; AAH08456.1; -. 2CF2AA7F1CDB707 CRC64;
SQ SEQUENCE 296 AA; 33428 MW; 2CF2AA7F1CDB707 CRC64;

Query Match 100.0%; Score 49; DB 4; Length 296;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
Db 289 QKYCCSRK 296
|||||||

RESULT 3
Q9KQZ2 PRELIMINARY; PRT; 703 AA.
ID Q9KQZ2 AC Q9KQZ2
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ATP-DEPENDENT HELICASE, DING FAMILY.
GN VC1855.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL: AE004261; AAF95003.1; -.
DR TIGR: VC1855; -.
KW Helicase; Complete proteome.
SQ SEQUENCE 703 AA; 78488 MW; FC88B361B8A8578C CRC64;

Query Match 77.6%; Score 38; DB 16; Length 703;
Best Local Similarity 62.5%; Pred. No. 17;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
Db 131 QKYCCSRK 138
|||||

RESULT 4
Q9JGP6 PRELIMINARY; PRT; 982 AA.
ID Q9JGP6 AC Q9JGP6
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OUTER CAPSID PROTEIN VP2.
GN L2.
OS Epizootic hemorrhagic disease virus.
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=10908;
[1]
```

```
RN SEQUENCE FROM N.A.
RX MEDLINE=20231376; PubMed=10770607;
RA Nara E., Iwata H., Inoue T.;
RT "The complete nucleotide sequence of segment L2 of Ibaraki virus
RT encoding for the antigen recognized by neutralizing antibodies.";
RL J. Vet. Med. Sci. 62:317-321(2000).
DR EMBL: AB030735; BAA93692.1; -.
DR InterPro: IPR001742; Orbi_VP2.
DR Pfam: PF00898; Orbi_VP2; 2.
DR ProDom: PD002938; Orbi_VP2; 1.
SQ SEQUENCE 982 AA; 115598 MW; 6E347B126158F775 CRC64;

Query Match 77.6%; Score 38; DB 12; Length 982;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
Db 951 EKFCGRK 958
|||||

RESULT 5
Q9UIS5 PRELIMINARY; PRT; 234 AA.
ID Q9UIS5 AC Q9UIS5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SMALL NUCLEAR RIBONUCLEOPROTEIN B.
GN SNRPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP Gray T.A., Martone D.L., Schaldach M.A., McCarrey J.R., Nicholls R.D.;
RT "Concerted Regulation and Molecular Evolution of the Duplicated SNRPB
RT and SNRPN loci.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF134825; AAD54488.1; -.
DR EMBL: AF134822; AAD54488.1; JOINED.
DR EMBL: AF134823; AAD54488.1; JOINED.
DR EMBL: AF134824; AAD54488.1; JOINED.
DR InterPro: IPR001163; snRNP_Sm.
DR Pfam: PF01423; Sm; 1.
KW Nucleocapsid; Ribonucleoprotein.
SQ SEQUENCE 234 AA; 26313 MW; A5898DB3C46A6E01 CRC64;

Query Match 75.5%; Score 37; DB 4; Length 234;
Best Local Similarity 71.4%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCSRK 7
Db 150 KRYCCSRK 156
|||||

RESULT 6
Q9UIS4 PRELIMINARY; PRT; 243 AA.
ID Q9UIS4 AC Q9UIS4
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SMALL NUCLEAR RIBONUCLEOPROTEIN B'.
GN SNRPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
```

```

RP SEQUENCE FROM N.A.
RA Gray T.A., Martone D.L., Schaldach M.A., McCarrey J.R., Nicholls R.D.;
RT "Concerted Regulation and Molecular Evolution of the Duplicated SNRPN
RL and SNRPN loci.";
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF134825; AAD54489.1; -.
DR EMBL; AF134822; AAD54489.1; JOINED.
DR EMBL; AF134823; AAD54489.1; JOINED.
DR EMBL; AF134824; AAD54489.1; JOINED.
DR InterPro; IPR001163; snRNP_Sm.
DR Pfam; PF01423; Sm; 1.
KW Nucleocapsid; Ribonucleoprotein.
SQ SEQUENCE 243 AA; 27267 MW; 05D3CAF973640CD5 CRC64;

Query Match 75.5%; Score 37; DB 4; Length 243;
Best Local Similarity 71.4%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCSR 7
Db :|||||
150 KRYCCSR 156

RESULT 7
O18233 PRELIMINARY; PRT; 571 AA.
AC O18233;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Y57G11C.7 PROTEIN.
GN Y57G11C.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z99281; CAB16509.1; -.
DR InterPro; IPR002900; DUF38.
DR InterPro; IPR001810; F-box.
DR Pfam; PF01827; DUF38; 2.
DR Pfam; PF00646; F-box; 1.
SQ SEQUENCE 571 AA; 66883 MW; DCB0A38465F67376 CRC64;

Query Match 75.5%; Score 37; DB 5; Length 571;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSR 8
Db :|||||
192 KRYCCPR 199

RESULT 8
Q9V8F4 PRELIMINARY; PRT; 58 AA.
AC Q9V8F4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG14503 PROTEIN.
GN CG14503.
OS Drosophila melanogaster (Fruit fly).

```

```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McPherson D.,
RA Arif J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.I., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Paizolo M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003800; AAF57713.1; -.
DR FlyBase; FBgn0040737; CG14503.
SQ SEQUENCE 58 AA; 6561 MW; 42ECFF3743C5930E CRC64;

Query Match 73.5%; Score 36; DB 5; Length 58;
Best Local Similarity 71.4%; Pred. No. 5.1;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 KYCCSR 8
Db :|||||
18 KYCCAER 24

RESULT 9
Q98S56 PRELIMINARY; PRT; 71 AA.
AC Q98S56;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE RNA POLYMERASE II SUBUNIT RPB10.
GN RPB10.
OS Guillardia theta (Cryptomonas phi).
OC Nucleomorph.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=21223349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.G.;
RA "The highly reduced genome of an enslaved algal nucleus.";
RT Nature 410:1091-1096(2001).
DR EMBL; AF083031; AAK39726.1; -.
DR InterPro; IPR000268; RNA_pol_N.
DR Pfam; PF01194; RNA_pol_N; 1.
SQ SEQUENCE 71 AA; 8391 MW; 7221827907FF5504 CRC64;

Query Match 73.5%; Score 36; DB 8; Length 71;
Best Local Similarity 71.4%; Pred. No. 6.1;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCSR 7
   | : | | | |
Db 41 QKYCRR 47

RESULT 10
Q15371 PRELIMINARY; PRT; 150 AA.
ID Q15371
AC Q15371;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RIBOSOMAL PROTEIN L18A HOMOLOGUE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zenz K.I.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X80821; CAA56787.1; -.
FT NON_TER 1
FT SEQUENCE 150 AA; 18050 MW; 4E5E933C97FA075F CRC64;

Query Match 73.5%; Score 36; DB 4; Length 150;
Best Local Similarity 75.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
   | : | | | |
Db 135 QKYCCLK 142

RESULT 11
O44161 PRELIMINARY; PRT; 263 AA.
ID O44161
AC O44161;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CD4.7 PROTEIN.
GN CD4.7
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,

```

```

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z., Scheet P.;
RT "The sequence of C. elegans cosmid CD4.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF036694; AAB88345.1; -.
SQ SEQUENCE 263 AA; 29908 MW; 9551D43D09B7BACD CRC64;

Query Match 73.5%; Score 36; DB 5; Length 263;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KYCCSR 7
   | : | | | |
Db 234 KYCCTR 239

RESULT 12
Q40142 PRELIMINARY; PRT; 88 AA.
ID Q40142
AC Q40142;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CELL WALL PROTEIN PRECURSOR.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. RUTGERS;
RX MEDLINE=95003699; PubMed=7919979;
RA Domingo C., Gomez M.D., Canas L., Hernandez-Yago J., Conejero V.,
RA Vera P.;
RT "A novel extracellular matrix protein from tomato plant associated
RT with lignified secondary cell wall.";
RL Plant Cell 6:1035-1047(1994).
DR EMBL; X77373; CAA54561.1; -.
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 88 CELL WALL PROTEIN.
SQ SEQUENCE 88 AA; 9919 MW; 128810E3B886284A CRC64;

Query Match 71.4%; Score 35; DB 10; Length 88;
Best Local Similarity 83.3%; Pred. No. 11;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCS 6
   | : | | | |
Db 72 KKYCCS 77

RESULT 13
O17778 PRELIMINARY; PRT; 127 AA.
ID O17778
AC O17778;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

```

DE F10D11.3 PROTEIN.
GN F10D11.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018 (1998).
DR EMBL; 281057; CAB02915.1; -.
SQ SEQUENCE 127 AA; 15008 MW; 6F015987A4E8446A CRC64;

Query Match 71.4%; Score 35; DB 5; Length 127;
Best Local Similarity 83.3%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 YCCSRK 8
:|||||
Db 118 FCCSRK 123

RESULT 14
ID Q9BK14 PRELIMINARY; PRT; 216 AA.
AC Q9BK14;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE THIOREDOXIN PEROXIDASE 2.
GN TPX2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=21153246; PubMed=11231293;
RA Rahfs S.; Becker K.;
RT "Thioredoxin peroxidases of the malarial parasite Plasmodium
RT falciparum."
RL Eur. J. Biochem. 268:1404-1409 (2001).
DR EMBL; AF225978; AAK20024.1; -.
KW Peroxidase.
SQ SEQUENCE 216 AA; 24680 MW; C73610CDD1015C1B CRC64;

Query Match 71.4%; Score 35; DB 5; Length 216;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKYCC 5
:|||||
Db 51 QKYCC 55

RESULT 15
Q9UGG9
ID Q9UGG9 PRELIMINARY; PRT; 330 AA.
AC Q9UGG9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BA422A16.1 (E1A BINDING PROTEIN P300) (FRAGMENT).
GN EP300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RA Blakey S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL096765; CAB62921.1; -.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS0014; BROMODOMAIN_2; 1.
FT NON_TER 1
FT NON_TER 330 330
SQ SEQUENCE 330 AA; 37707 MW; 01F502BCBFA7C751 CRC64;

Query Match 71.4%; Score 35; DB 4; Length 330;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 YCCSRK 8
:|||||
Db 223 YCCGRK 228

Search completed: November 1, 2002, 14:49:53
Job time : 4.45338 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:41:32 ; Search time 0.411576 Seconds
(without alignments)
752.611 Million cell updates/sec

Title: US-09-606-129A-35
Perfect score: 49
Sequence: 1 QKYCCSRK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	49	100.0	296	1 BIEA_HUMAN	P53004 homo sapien
2	39	79.6	963	1 ATC4_HUMAN	O75185 homo sapien
3	38	77.6	151	1 YOM1_PROPR	P29741 photobacter
4	36	73.5	71	1 RPBX_SCHPO	O13877 schizosacch
5	36	73.5	761	1 TOPI_MERJA	Q59046 methanococc
6	35	71.4	73	1 RPNL_METJA	Q57649 methanococc
7	35	71.4	202	1 T4S1_MOUSE	Q64302 mus musculu
8	35	71.4	652	1 ABG5_MOUSE	Q99pe8 mus musculu
9	35	71.4	709	1 ABG5_RAT	Q99pe7 rattus norv
10	35	71.4	709	1 SWI5_YEAST	P08153 saccharomyc
11	35	71.4	859	1 YDBB_SCHPO	Q10362 schizosacch
12	35	71.4	2414	1 P300_HUMAN	Q09472 homo sapien
13	35	71.4	2441	1 CBP_MOUSE	P45481 mus musculu
14	35	71.4	2442	1 CBP_HUMAN	Q92793 homo sapien
15	34	69.4	37	1 TXOF_HADVE	P81599 hadronyche
16	34	69.4	73	1 MPT5_AMBTR	P10414 ambrosia tr
17	34	69.4	76	1 SPY1_CHICK	Q9pt11 gallus gall
18	34	69.4	139	1 SPY1_HUMAN	Q43609 homo sapien
19	34	69.4	313	1 SPY1_MOUSE	Q9qvx9 mus musculu
20	34	69.4	418	1 EPF2_RALSO	Q45412 ralstonia s
21	34	69.4	503	1 YMJ6_CAEEL	P34482 caenorhabdi
22	34	69.4	1691	1 POLN_HEVME	Q03495 hepatitis e
23	34	69.4	1693	1 POLN_HEVBU	Q03324 hepatitis e
24	34	69.4	1693	1 POLN_HEVNY	Q04610 hepatitis e
25	34	69.4	1693	1 POLN_HEVPA	P33424 hepatitis e
26	34	69.4	4543	1 LRPI_CHICK	P98157 gallus gall
27	33	67.3	65	1 RPNL_PYRAB	Q9v194 pyrococcus
28	33	67.3	65	1 RPNL_PYRHO	O59298 pyrococcus
29	33	67.3	72	1 RPNL_THEAC	Q9h109 thermoplasma
30	33	67.3	76	1 RPNL_ARCFU	O29135 archaeoglob
31	33	67.3	98	1 RK23_ASTIO	P34771 astasia lon
32	33	67.3	183	1 YJRQ_YEAST	P46993 saccharomyc
33	33	67.3	210	1 YRBL_ECOLI	P46021 escherichia

34	33	67.3	238	1 TPML_LYCES	O01591 lycopersico
35	33	67.3	246	1 OS13_SOLCO	P50701 solanum com
36	33	67.3	246	1 OSMO_TOBAC	P14170 nicotiana t
37	33	67.3	247	1 NP24_LYCES	P12670 lycopersico
38	33	67.3	247	1 OS81_SOLCO	P50702 solanum com
39	33	67.3	250	1 OS35_SOLCO	P50703 solanum com
40	33	67.3	251	1 OLPA_TOBAC	P25871 nicotiana t
41	33	67.3	2045	1 CBP1_CAEEL	P34545 caenorhabdi
42	32	65.3	27	1 CXK7_CONPU	P56633 conus purpu
43	32	65.3	55	1 RPNL_MEYTH	Q36147 sulfolobus
44	32	65.3	66	1 RPNL_SULSO	Q98058 sulfolobus
45	32	65.3	67	1 RPBX_DROME	Q9vc49 drosophila

ALIGNMENTS

RESULT 1					
BIEA_HUMAN	BIEA_HUMAN	STANDARD;	PRT;	296 AA.	
AC	P53004;	1996 (Rel. 34, Created)			
DT	01-OCT-1996	(Rel. 34, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Biliverdin reductase A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-reductase).				
DE					
GN	BLVRA OR BLVR OR BVR.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Placenta;				
RX	MEDLINE=96202961; PubMed=8631357;				
RA	Maines M.D., Polevoda B.V., Huang T.-J., McCoubrey W.K. Jr.;				
RT	"Human biliverdin IXalpha reductase is a zinc-metalloprotein.				
RT	Characterization of purified and Escherichia coli expressed				
RT	enzymes.";				
RL	Eur. J. Biochem. 235:372-381(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Komuro A., Tobe T., Nakano Y., Yamaguchi T., Tomita M.;				
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 1-117 FROM N.A.				
RA	Cordes M., Wollam C., Carter T.;				
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	SEQUENCE OF 3-36; 48-74 AND 228-248.				
RC	TISSUE=Liver;				
RX	MEDLINE=93143333; PubMed=8424666;				
RA	Maines M.D., Trakshel G.M.;				
RT	"Purification and characterization of human biliverdin reductase.";				
RL	Arch. Biochem. Biophys. 300:320-326(1993).				
RN	[5]				
RP	SEQUENCE OF 3-22.				
RC	TISSUE=Liver;				
RX	MEDLINE=95014177; PubMed=7929092;				
RA	Yamaguchi T., Komoda Y., Nakajima H.;				
RT	"Biliverdin-IX alpha reductase and biliverdin-IX beta reductase from				
RT	human liver. Purification and characterization.";				
RL	J. Biol. Chem. 269:24343-24348(1994).				
CC	- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN.				
CC	- COFACTORS: BILIRUBIN + NAD(P)(+) = biliverdin + NAD(P)H.				
CC	- COFACTORS: BINDS ONE ZINC ION. HAS DUAL PH/COFACTOR (NADH, NADPH).				
CC	- SPECIFICITY: USES NADH AT THE ACIDIC PH RANGE (6-6.7) AND NADPH AT				
CC	THE ALKALINE RANGE (8.5-8.7).				
CC	- PATHWAY: FINAL STEP IN HEME METABOLISM.				
CC	- SUBUNIT: MONOMER.				
CC	- SUBCELLULAR LOCATION: Cytoplasmic.				
CC	- TISSUE SPECIFICITY: LIVER.				
CC	- SIMILARITY: TO E.COLI YHHX.				

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; X93086; CAA63635.1; -;
EMBL; U34877; AAC35588.1; -;
EMBL; AAC005189; AAC25526.1; -;
MIM; 109750; -;
InterPro: IPR000683; GFO_IDH_MOCA.
Pfam: PF01408; GFO_IDH_MOCA. 1.
Oxidoreductase; NAD; NADP; Zinc.
PROPEP 1 2
CHAIN 3 296 BILIVERDIN REDUCTASE A.
FT DOMAIN 11 16 POLY-VAL.
FT METAL 280 280 ZINC (POTENTIAL).
FT METAL 281 281 ZINC (POTENTIAL).
FT METAL 292 292 ZINC (POTENTIAL).
FT METAL 293 293 ZINC (POTENTIAL).
FT CONFLICT 3 3 A -> T (IN REF. 2).
FT CONFLICT 154 155 SD -> AG (IN REF. 2).
FT CONFLICT 160 160 D -> E (IN REF. 2).
FT SEQUENCE 296 AA; 33488 MW; 0DFD3B386F4DFC0A CRC64;

Query Match 100.0%; Score 49; DB 1; Length 296;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
|||||||
DB 289 QKYCCSRK 296

RESULT 2
ATC4_HUMAN STANDARD; PRT; 963 AA.
AC 075185; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable calcium-transporting ATPase KIAA0703 (EC 3.6.3.8).
GN KIAA0703
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:169-176(1998).
CC -1- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS
CC OF ATP COUPLED WITH THE TRANSPORT OF CALCIUM (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + CA(2+)(IN) = ADP + PHOSPHATE +
CC CA(2+)(OUT).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES). SUBFAMILY IIA.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; AB014603; BAA31678.1; -;
HSP; P04191; IEUL.
InterPro: IPR004014; Cation_ATPase.
InterPro: IPR001757; E1-E2_ATPase.
InterPro: IPR001454; Hydrolase.
InterPro: IPR000661; Na_H_K_ATPase.
Pfam: PF00689; Cation_ATPase_C; 1.
Pfam: PF00690; Cation_ATPase_N; 1.
Pfam: PF00122; E1-E2_ATPase; 1.
Pfam: PF00702; Hydrolase; 1.
PRINTS: PR00119; CATATPASE.
PROSITE: PS00154; ATPASE_E1_E2; 1.
Hydrolase; Calcium transport; Transmembrane; Phosphorylation;
KW ATP-binding; Metal-binding; Magnesium; Calcium-binding;
KW Multigene family.
FT DOMAIN 1 118 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 119 139 1 (BY SIMILARITY).
FT DOMAIN 140 151 LUMENAL (BY SIMILARITY).
FT TRANSMEM 152 170 2 (BY SIMILARITY).
FT DOMAIN 171 308 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 309 328 3 (BY SIMILARITY).
FT DOMAIN 329 340 LUMENAL (BY SIMILARITY).
FT TRANSMEM 341 358 4 (BY SIMILARITY).
FT DOMAIN 359 746 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 747 766 5 (BY SIMILARITY).
FT DOMAIN 767 776 LUMENAL (BY SIMILARITY).
FT TRANSMEM 777 797 6 (BY SIMILARITY).
FT DOMAIN 798 817 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 818 840 7 (BY SIMILARITY).
FT DOMAIN 841 856 LUMENAL (BY SIMILARITY).
FT TRANSMEM 857 876 8 (BY SIMILARITY).
FT DOMAIN 877 889 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 890 908 LUMENAL (BY SIMILARITY).
FT DOMAIN 909 923 9 (BY SIMILARITY).
FT TRANSMEM 924 944 10 (BY SIMILARITY).
FT DOMAIN 945 963 CYTOPLASMIC (BY SIMILARITY).
FT MOD_RES 963 963 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 991 991 MAGNESIUM (BY SIMILARITY).
FT METAL 995 995 MAGNESIUM (BY SIMILARITY).
FT CA_BIND 349 349 SITE 2 (BY SIMILARITY).
FT CA_BIND 350 350 SITE 2 (BY SIMILARITY).
FT CA_BIND 352 352 SITE 2 (BY SIMILARITY).
FT CA_BIND 354 354 SITE 2 (BY SIMILARITY).
FT CA_BIND 785 785 SITE 2 (BY SIMILARITY).
FT CA_BIND 789 789 SITE 2 (BY SIMILARITY).
FT SEQUENCE 963 AA; 105036 MW; C8CCE7BE29C43589 CRC64;

Query Match 79.6%; Score 39; DB 1; Length 963;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
|||||
DB 947 QKYCCSPK 954

RESULT 3
YOM1_PHOPR STANDARD; PRT; 151 AA.
ID YOM1_PHOPR
AC P29741;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ATP-dependent helicase in ompH 5' region (ORF1) (Fragment).
OS Photobacterium profundum (Photobacterium sp. (strain SS9)).
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae;
OC Photobacterium.
OX NCBI_TaxID=74109;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=93380661; PubMed=8396546;
RA Bartlett D.H., Chi E., Wright M.E.;
RT "Sequence of the omph gene from the deep-sea bacterium Photobacterium
RL S29.";
RN Gene 131:125-128(1993).
RP
RX SIMILARITY TO DING.
RA MEDLINE=93219138; PubMed=8385320;
RA Koonin E.V.;
RT "Escherichia coli ding gene encodes a putative DNA helicase related
RL to a group of eukaryotic helicases including Rad3 protein.";
RN Nucleic Acids Res. 21:1497-1497(1993).
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. DING SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67094; CAA47468.1; -
DR PIR; S23215; S23215.
KW Hypothetical protein; Hydrolase; Helicase; ATP-binding;
KW DNA-binding.
FT NP_BIND 56 63 ATP (PROBABLE).
FT NON_TER 151 151
SQ SEQUENCE 151 AA; 17086 MW; FC8FA7862EAC1805 CRC64;

Query Match 77.6%; Score 38; DB 1; Length 151;
Best Local Similarity 62.5%; Pred. No. 6.6;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKYCCSRK 8
Db 119 QRYCCNHK 126
I:||||:|

RESULT 4
RPBX_SCHPO STANDARD; PRT; 71 AA.
AC O13877; O14458;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerases I, II, and III 8.3 kDa polypeptide
DE (EC 2.7.7.6) (ABC10-beta).
GN RPB10 OR SPAC1B3.12C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA MEDLINE=97177445; PubMed=9054344;
RA Shpakovskii G.V., Proshkin S.A., Lebedenko E.N.;
RT "Three regions of Rpb10 mini-subunit of nuclear RNA polymerases are
RL strictly conserved in all eukaryotes (letter).";
RN Bloorg. Khim. 22:938-940(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=98271918; PubMed=9608944;
RA Shpakovskii G.V., Proshkin S.A., Lebedenko E.N.;
RT "Exon-intron organization rpb10+ and rpb10+ genes of
RN Schizosaccharomyces pombe, coding for mini-subunits of nuclear RNA-
RN polymerase I-III.";
RN Mol. Biol. (Mosk) 32:285-290(1998).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 15
CC DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS SHARED BY ALL 3 YEAST RNA
CC POLYMERASES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPB10 / EUKARYOTIC RPB10
CC RNA POLYMERASE SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X95733; CAA65049.1; -
DR EMBL; AF027818; AAC16895.1; -
DR EMBL; D89596; BAA22805.1; -
DR EMBL; U80219; AAC49842.1; -
DR EMBL; 298598; CAB11246.1; -
DR HSSP; O26147; 1EF4.
DR InterPro; IPR002068; RNA_pol_N.
DR Pfam; PF01194; RNA_pol_N; 1.
DR PROSITE; PS01112; RNA_POL_N_8KD; 1.
KW Transferrase; DNA-directed RNA polymerase; Transcription;
KW Nuclear protein; Zinc; Metal-binding.
FT METAL 7 7 ZINC (BY SIMILARITY).
FT METAL 10 10 ZINC (BY SIMILARITY).
FT METAL 44 44 ZINC (BY SIMILARITY).
FT METAL 45 45 ZINC (BY SIMILARITY).
SQ SEQUENCE 71 AA; 8276 MW; 8B9EE8FD59A2458 CRC64;

Query Match 73.5%; Score 36; DB 1; Length 71;
Best Local Similarity 71.4%; Pred. No. 7.7;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKYCCSR 7
Db 41 QRYCCRR 47
I:||||:|

RESULT 5
TOP1_METJA STANDARD; PRT; 761 AA.
AC Q59046;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Untwisting enzyme) (Swivelase).
GN TOPA OR MJ1652.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,

```

RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
 CC DNA, followed by passage and rejoining.
 CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK IN
 CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
 CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67605; AAB99673.1; -;
 DR HSSP; P06612; IECL.
 DR TIGR; MJ1652; -;
 DR InterPro; IPR003601; DNATopI_ATP_bind.
 DR InterPro; IPR003602; DNATopI_DNA_bind.
 DR InterPro; IPR000380; Pro_topoisomerase.
 DR InterPro; IPR002936; Toprim.
 DR Pfam; PF01131; Topoisom_bac; 1.
 DR Pfam; PF01751; Toprim; 1.
 DR Pfam; PF01396; zf-C4_Topoiso; 3.
 DR PRINTS; PR00417; PRTPISMRASEI.
 DR SMART; SM00437; TOP1AC; 1.
 DR SMART; SM00436; TOP1BC; 1.
 DR SMART; SM00493; TOPRIM; 1.
 DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
 DR Isomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;
 KW Repeat; Complete proteome.
 FT ZN_FING 600 626 C4-TYPE 1.
 FT ZN_FING 680 706 C4-TYPE 2.
 FT ZN_FING 721 747 C4-TYPE 3.
 FT ACT_SITE 315 315 DNA_CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 761 AA; 87832 MW; 2F9C95753E202D82 CRC64;
 Query Match 73.5%; Score 36; DB 1; Length 761;
 Best Local Similarity 71.4%; Pred. No. 49;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 KYCCSRK 8
 Db 125 KYCCGRE 131
 RESULT 6
 RPON_METJA
 ID RPON_METJA STANDARD; PRT; 73 AA.
 AC Q57649;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).
 GN RPON OR MJ0196
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.

OX NCBI_TaxID=2190;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RC MEDLINE=96337999; PubMed=8688087;
 RX Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Hanna M.C.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Borodovsky M.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -1- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPON / EUKARYOTIC RPBI0
 CC RNA POLYMERASE SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67475; AAB98176.1; -;
 DR HSSP; O26147; 1EP4.
 DR TIGR; MJ0196; -;
 DR InterPro; IPR000268; RNA_pol_N.
 DR Pfam; PF01194; RNA_pol_N; 1.
 DR PROSITE; PS01112; RNA_POL_N_8KD; 1.
 KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Metal-binding; Complete proteome.
 FT METAL 7 7 ZINC (BY SIMILARITY).
 FT METAL 10 10 ZINC (BY SIMILARITY).
 FT METAL 44 44 ZINC (BY SIMILARITY).
 FT METAL 45 45 ZINC (BY SIMILARITY).
 SQ SEQUENCE 73 AA; 8695 MW; E716EA406D65B831 CRC64;
 Query Match 71.4%; Score 35; DB 1; Length 73;
 Best Local Similarity 71.4%; Pred. No. 11;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 QKYCCSR 7
 Db 41 KKYCCRR 47
 RESULT 7
 T4SL_MOUSE
 ID T4SL_MOUSE STANDARD; PRT; 202 AA.
 AC Q64302;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transmembrane 4 superfamily, member 1 (Tumor-associated antigen L6)
 DE (Membrane component, surface marker 1) (M3S1).
 GN TM4SF1 OR M3S1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND TOPOLOGY.
 RC STRAIN=BALB/C; TISSUE=Thymic epithelium;
 RX MEDLINE=94171760; PubMed=7510285;

RA Marken J.S., Bajorath J., Edwards C.P., Farr A.G., Schieven G.L.,
RA Hellstroem I., Hellstroem K.E., Aruffo A.;
RT "Membrane topology of the L6 antigen and identification of the protein
RL J. Biol. Chem. 269:7397-7401(1994).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SKIN AND LUNG. MODERATELY
CC EXPRESSED IN LYMPH NODES AND KIDNEYS. ALSO PRESENT IN THYMIC
CC STROMA AND FIBROBLASTS.
CC -1- SIMILARITY: BELONGS TO THE TETRASPANIN (TM4SF) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: LI5443; AAA39403.1; -;
CC EMBL: LI5429; AAA17873.1; -;
CC MGD: MGI:104678; Tm4sf1.
CC Glycoprotein; Antigen; Transmembrane.
CC DOMAIN 1 9 CYTOPLASMIC (PROBABLE).
CC FT TRANSSEM 10 30 PROBABLE.
CC FT DOMAIN 31 49 EXTRACELLULAR (PROBABLE).
CC FT TRANSSEM 50 70 PROBABLE.
CC FT DOMAIN 71 93 CYTOPLASMIC (PROBABLE).
CC FT TRANSSEM 94 114 PROBABLE.
CC FT DOMAIN 115 161 EXTRACELLULAR (PROBABLE).
CC FT TRANSSEM 162 182 PROBABLE.
CC FT DOMAIN 183 202 CYTOPLASMIC (PROBABLE).
CC FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 202 AA; 22243 MW; 93839E707D85FF02 CRC64;

Query Match 71.4%; Score 35; DB 1; Length 202;
Best Local Similarity 83.3%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YCCSRK 8
| | | | |
Db 192 YCCSRQ 197

RESULT 8
ABG5_MOUSE
ID ABG5_MOUSE STANDARD; PRT; 652 AA.
AC Q99PE8;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABCG5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=20578753; PubMed=11138003;
RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
RA Dean M., Patel S.B.;
RT "Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption.";
RL Nat. Genet. 27:79-83(2001).
RN [2]
RP TISSUE SPECIFICITY, AND INDUCTION.
RX MEDLINE=2053648; PubMed=11099417;
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RA Kwitterovich P., Shan B., Barnes R., Hobbs H.H.;

RT "Accumulation of dietary cholesterol in sitosterolemia caused by
RT mutations in adjacent ABC transporters.";
RL Science 290:1771-1775(2000).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion (by similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
CC level, in the liver.
CC -1- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF312713; AAC53097.1; -;
CC MGD: MGI:1351659; Abcg5.
CC InterPro: IPR003593; AAA.
CC InterPro: IPR003439; ABC-transportr.
CC InterPro: IPR001687; ATP-GTP_A.
CC Pfam: PF00005; ABC_tran; 1.
CC SMART: SM00382; AAA; 1.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC ATP-binding: Glycoprotein; Transmembrane; Transport.
CC DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
CC FT TRANSSEM 386 406 1 (POTENTIAL).
CC FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
CC FT TRANSSEM 423 443 2 (POTENTIAL).
CC FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
CC FT TRANSSEM 464 484 3 (POTENTIAL).
CC FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
CC FT TRANSSEM 505 525 4 (POTENTIAL).
CC FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
CC FT TRANSSEM 530 550 5 (POTENTIAL).
CC FT DOMAIN 551 622 EXTRACELLULAR (POTENTIAL).
CC FT TRANSSEM 623 643 6 (POTENTIAL).
CC FT DOMAIN 644 652 CYTOPLASMIC (POTENTIAL).
CC FT NP_BIND 87 94 ATP (POTENTIAL).
CC FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match 71.4%; Score 35; DB 1; Length 652;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCC 5
| | | | |
Db 569 QKYCC 573

RESULT 9
ABG5_RAT
ID ABG5_RAT STANDARD; PRT; 652 AA.
AC Q99PE7;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABCG5.
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=101116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Small intestine;
 RX MEDLINE=20578753; PubMed=11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 RA Dean M., Patel S.B.;
 RT "Identification of a gene, ABCG5, important in the regulation of
 RT dietary cholesterol absorption."
 RL Nat. Genet. 27:79-83(2001).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF312714; AAG53098.2; -;
 DR InterPro: IPR003593; AAA;
 DR InterPro: IPR003439; ABC_transportr.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 385
 FT TRANSMEM 386 406
 FT DOMAIN 407 422
 FT TRANSMEM 423 443
 FT DOMAIN 444 463
 FT TRANSMEM 464 484
 FT DOMAIN 485 504
 FT TRANSMEM 505 525
 FT DOMAIN 526 529
 FT TRANSMEM 530 550
 FT DOMAIN 551 624
 FT TRANSMEM 625 645
 FT DOMAIN 646 652
 FT NP_BIND 87 94
 FT CARBOHYD 585 585
 FT CARBOHYD 592 592
 SQ SEQUENCE 652 AA; 73342 MW; 4D42FE2BAB0DAD59 CRC64;
 Query Match 71.4%; Score 35; DB 1; Length 652;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QKYCC 5
 Db 569 QKYCC 573
 RESULT 10
 ID SWI5_YEAST
 AC P08153;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcriptional factor SWI5.
 GN SWI5 OR YDR146C OR YD8358.03C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 ON NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88211561; PubMed=3284746;
 RA Stillman D.J., Bankier A.T., Seddon A., Groenhout E.G., Nasmyth K.A.;
 RT "Characterization of a transcription factor involved in mother cell
 RT specific transcription of the yeast HO gene."
 RL EMBO J. 7:485-494(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Murphy L., Richards C., Harris D., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP DNA-BINDING.
 RX MEDLINE=88156968; PubMed=2831463;
 RA Nagai K., Nakaseko Y., Nasmyth K.A., Rhodes D.;
 RT "Zinc-finger motifs expressed in E. coli and folded in vitro direct
 RT specific binding to DNA."
 RL Nature 332:284-286(1988).
 RN [4]
 RP PHOSPHORYLATION, MUTAGENESIS, AND INTRACELLULAR LOCALIZATION.
 RX MEDLINE=91347374; PubMed=1652372;
 RA Moll T., Tebb G., Surana U., Robitsch H., Nasmyth K.;
 RT "The role of phosphorylation and the CDC28 protein kinase in cell
 RT cycle-regulated nuclear import of the S. cerevisiae transcription
 RT factor SWI5."
 RL Cell 66:743-758(1991).
 RN [5]
 RP STRUCTURE BY NMR OF 540-608.
 RX MEDLINE=93085741; PubMed=1453468;
 RA Neuhaus D., Nakaseko Y., Schwabe J.W.R., Klug A.;
 RT "Solution structures of two zinc-finger domains from SWI5 obtained
 RT using two-dimensional 1H nuclear magnetic resonance spectroscopy. A
 RT zinc-finger structure with a third strand of beta-sheet."
 RL J. Mol. Biol. 228:637-651(1992).
 CC -!- FUNCTION: DETERMINES THE MOTHER-CELL-SPECIFIC TRANSCRIPTION OF THE
 CC HO ENDONUCLEASE GENE THAT IS RESPONSIBLE FOR THE INITIATION OF
 CC MATING-TYPE SWITCHING IN YEAST. RECOGNIZES A SPECIFIC SEQUENCE IN
 CC THE PROMOTER OF THE HO GENE. ACTIVATES EGT2 TRANSCRIPTION IN A
 CC CONCENTRATION-DEPENDENT MANNER. SYNTHESIZED DURING G2 AND EARLY
 CC MITOSIS.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR IN G1, BUT CYTOPLASMIC IN S, G2, AND
 CC M CELL CYCLE PHASES.
 CC -!- PTM: CELL CYCLE-DEPENDENT PHOSPHORYLATION OF THREE SERINE RESIDUES
 CC PREVENTS SWI5 FROM ENTERING THE NUCLEUS, AND IT ACCUMULATES IN THE
 CC CYTOPLASM. AS A CONSEQUENCE OF CDC28 KINASE INACTIVATION AT THE
 CC END OF ANAPHASE, THE THREE SERINE RESIDUES ARE DEPHOSPHORYLATED
 CC AND SWI5 ENTERS THE NUCLEUS TO ACTIVATE TRANSCRIPTION. IT IS THEN
 CC RAPIDLY DEGRADED. THREONINE PHOSPHORYLATION ALSO SEEMS TO OCCUR.
 CC -!- SIMILARITY: STRONG, TO YEAST METALLOTHIONEIN EXPRESSION ACTIVATOR
 CC ACE2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06978; CAA30040.1; -;
 DR EMBL; 250046; CAA90369.1; -;
 DR PIR; S00342; TWBY55.
 DR PDB; 1NCS; 10-JUL-96.
 DR PDB; 1ZFD; 14-OCT-96.

DR TRANSFAC; T00776; -
DR SGD; S0002553; SWIS.
DR InterPro; IPR000637; AT_hoek.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF02178; AT_hoek; 1.
DR Pfam; PF00096; zf-C2H2; 3.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00384; AT_hoek; 1.
DR SMART; SM00355; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 2.
KW Transcription regulation; Activator; Zinc-finger; DNA-binding;
KW Repeat; Metal-binding; Nuclear protein; Phosphorylation;
KW 3D-structure.
FT DOMAIN 550 632 ZINC FINGERS.
FT ZN_FING 550 574 C2H2-TYPE.
FT ZN_FING 580 604 C2H2-TYPE.
FT ZN_FING 609 632 C2H2-TYPE.
FT SITE 635 659 NUCLEAR TARGETING SIGNAL (POTENTIAL).
FT MOD_RES 522 522 PHOSPHORYLATION (BY CDC28).
FT MOD_RES 646 646 PHOSPHORYLATION (BY CDC28).
FT MOD_RES 664 664 PHOSPHORYLATION (BY CDC28).
FT MUTAGEN 522 522 S->A: CONSTITUTIVE NUCLEAR ENTRY; WHEN
FT MUTAGEN 646 646 ASSOCIATED WITH A-646 AND A-664.
FT MUTAGEN 664 664 S->A: CONSTITUTIVE NUCLEAR ENTRY; WHEN
FT MUTAGEN 664 664 ASSOCIATED WITH A-522 AND A-664.
FT SEQUENCE 709 AA; 79775 MW; BEF5ED5BF6E30F6 CRC64;
Query Match 71.4%; Score 35; DB 1; Length 709;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative. 0; Mismatches 2; Indels 0; Gaps 0;
Qy . 1 QKYCCSRK 8
Db 269 QKYCLQK 276
RESULT 11
ID YDBB_SCHPO STANDARD; PRT; 859 AA.
AC Q10362;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 94.9 kDa protein C22E12.11c in chromosome I.
GN SPAC22E12.11c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Devlin K., Church C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -!- SIMILARITY: TO YEAST YKR029C AND YJL105W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z70043; CAA93898.1; -
DR InterPro; IPR001965; PHD.
DR InterPro; IPR001214; SET.

DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS0280; SET; 1.
KW Hypothetical protein; Zinc-finger.
FT DOMAIN 226 341 SET.
SQ SEQUENCE 859 AA; 94886 MW; B4865BF40FD2C5D8 CRC64;
Query Match 71.4%; Score 35; DB 1; Length 859;
Best Local Similarity 83.3%; Pred. No. 77;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 2 KYCCSR 7
Db 221 KFCCSR 226
RESULT 12
ID P300_HUMAN STANDARD; PRT; 2414 AA.
AC Q09472;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE E1A-associated protein p300.
GN EP300 OR P300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95011587; PubMed=7523245;
RA Eckner R., Ewen M.E., Newsome D., Gerdes M., Decaprio J.A.,
RA Lawrence J.B., Livingston D.M.;
RT "Molecular cloning and functional analysis of the adenovirus E1A-
RT associated 300-kD protein (p300) reveals a protein with properties of
RT a transcriptional adaptor.";
RL Genes Dev. 8:869-884(1994).
CC -!- FUNCTION: PROBABLE TRANSCRIPTIONAL ADAPTOR REQUIRED FOR THE
CC ACTIVITY OF CERTAIN COMPLEX TRANSCRIPTIONAL REGULATORY ELEMENTS.
CC MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO AND MAY BE
CC INVOLVED IN THE TRANSFORMING CAPACITY OF THE ADENOVIRUS E1A
CC PROTEIN.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZZ-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U01877; AAA18639.1; -
DR TRANSFAC; T01427; -
DR MIN; 602700; -
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR003101; KIX.
DR InterPro; IPR000197; TAZ_finger.
DR InterPro; IPR000433; Znf_ZZ.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02172; KIX; 1.
DR Pfam; PF02135; zf-TAZ; 2.
DR Pfam; PF00569; ZZ; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00291; Znf_ZZ; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.

```
DR PROSITE; PS01357; ZF_ZZ_1; 1.
DR PROSITE; PS0135; ZF_ZZ_2; 1.
KW Transcription regulation; Nuclear protein; Bromodomain; Cell cycle;
KW Zinc-finger.
FT DOMAIN 11 17 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1067 1139 BROMODOMAIN.
FT DOMAIN 1572 1818 BINDING REGION FOR E1A ADENOVIRUS.
FT ZN_FING 1664 1707 ZZ-TYPE.
FT DOMAIN 797 800 POLY-SER.
FT DOMAIN 1519 1526 POLY-GLU.
FT DOMAIN 2066 2069 POLY-GLN.
FT DOMAIN 2190 2195 POLY-GLN.
SQ SEQUENCE 2414 AA; 264143 MW; 6BF909EE4B9D693 CRC64;

Query Match 71.4%; Score 35; DB 1; Length 2414;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 YCCSRK 8
DB 1162 YCCGRK 1167

RESULT 13
CBP_MOUSE
ID CBP_MOUSE STANDARD; PRT; 2441 AA.
AC P45481;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94019866; PubMed=8413673;
RA Chivlia J.C., Kwok R.P.S., Lamb N., Hagiwara M., Montminy M.R.,
RA Goodman R.H.;
RT "Phosphorylated CREB binds specifically to the nuclear protein CBP.";
RL Nature 365:855-859(1993).
CC -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZZ-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S66385; AAB28651.1; -.
CC TRANSFAC; T01318; -.
CC MGD; MGI:1098280; Crebbp.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR003101; KIX.
CC InterPro; IPR000197; TAZ_finger.
CC InterPro; IPR000433; Znf_ZZ.
CC Pfam; PF00439; bromodomain; 1.
CC Pfam; PF02172; KIX; 1.
CC Pfam; PF02135; zf-TAZ; 2.
CC Pfam; PF00569; ZZ; 1.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00297; BROMO; 1.
```

```
DR SMART; SM00291; Znf_ZZ; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01357; ZF_ZZ_1; 1.
DR PROSITE; PS0135; ZF_ZZ_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Zinc-finger.
FT DOMAIN 1104 1176 BROMODOMAIN.
FT ZN_FING 1702 1745 ZZ-TYPE.
FT DOMAIN 1062 1065 POLY-GLU.
FT DOMAIN 1556 1563 POLY-GLU.
FT DOMAIN 1944 1949 POLY-PRO.
FT DOMAIN 1968 1971 POLY-GLN.
FT DOMAIN 2082 2086 POLY-GLN.
FT DOMAIN 2200 2216 POLY-GLN.
FT DOMAIN 2296 2299 POLY-GLN.
SQ SEQUENCE 2441 AA; 265474 MW; 0ABB028C3112F419 CRC64;

Query Match 71.4%; Score 35; DB 1; Length 2441;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 YCCSRK 8
DB 1199 YCCGRK 1204

RESULT 14
CBP_HUMAN
ID CBP_HUMAN STANDARD; PRT; 2442 AA.
AC Q92793; Q16376; O00147;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97385172; PubMed=9238046;
RA Sobulo O.M., Borow J., Tomek R., Reshmi S., Harden A.,
RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
RA Zeleznik-Le N.J.;
RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=97321049; PubMed=9177780;
RA Giles R.H., Petrif F., Dauwerse H.G., den Hollander A.I.,
RA Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA Doggett N.A., Peters D.J.M., Breuning M.H.;
RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis
RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT 16p13.3.";
RL Genomics 42:96-144(1997).
RN [3]
RN SEQUENCE OF 1-405 FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Clivin C.I., Distche C., Dube I., Frischauf A.M.,
RA Housman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -!- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC T(8;16)(P11;P13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)
CC INVOLVING CBP AND MLL.
CC -1- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
CC SYNDROME (RBS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN
CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U47741; AAC51770.1; -;
DR EMBL: U85962; AAC51331.1; -;
DR EMBL: U89354; AAC51339.1; -;
DR EMBL: U89355; AAC51340.1; -;
DR MIM: 600140; -;
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003101; KIX.
DR InterPro: IPR000197; TAZ_finger.
DR InterPro: IPR000433; ZnF_ZZ.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF02172; KIX; 1.
DR Pfam: PF02135; zf-TAZ; 2.
DR Pfam: PF00569; ZZ; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00291; ZnF_ZZ; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS00014; BROMODOMAIN_2; 1.
DR PROSITE: PS01357; ZF_ZZ_1; 1.
DR PROSITE: PS0135; ZF_ZZ_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Chromosomal translocation; Zinc-finger.
FT ZN_FING 1701 1744
FT DOMAIN 363 430
FT DOMAIN 452 683
FT DOMAIN 1103 1175
FT DOMAIN 1061 1064
FT DOMAIN 1199 1487
FT DOMAIN 1555 1562
FT DOMAIN 1675 1849
FT DOMAIN 1943 1948
FT DOMAIN 1967 1970
FT DOMAIN 2081 2085
FT DOMAIN 2199 2216
FT DOMAIN 2245 2248
FT DOMAIN 2297 2300
FT DOMAIN 1511 1513
FT CONFLICT 1724 1725
FT CONFLICT 1770 1770
FT CONFLICT 1789 1789
FT CONFLICT 1812 1812
SQ SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;

Query Match 71.4%; Score 35; DB 1; Length 2442;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 YCCSRK 8
Db 1198 YCCGRK 1203

RESULT 15
TXOF_HADVE

ID TXOF_HADVE STANDARD; PRT; 37 AA.
AC P81599;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
OS Omega-atracotoxin-Hv1f (Omega-AcTx-Hv1f).
OS Hadronyche versuta (Blue mountains funnel-web spider) (Atrax
versutus).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Hexathelidae; Hadronyche.
OX NCBI_TaxID=6904;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=99421654; PubMed=10491095;
RA Wang X.-H., Smith R., Fletcher J.I., Wilson H., Wood C.J.,
RA Meriin E.H., King G.F.;
RT "Structure-function studies of w-atracotoxin, a potent antagonist
of insect voltage-gated calcium channels.";
RL Eur. J. Biochem. 264:488-494(1999).
CC -1- FUNCTION: INHIBITS INSECT, BUT NOT MAMMALIAN, VOLTAGE-GATED
CC CALCIUM CHANNEL CURRENTS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC HSP; P56207; IAXH.
KW Calcium channel inhibitor; Toxin; Neurotoxin.
FT DISULFID 4 18 BY SIMILARITY.
FT DISULFID 11 22 BY SIMILARITY.
FT DISULFID 17 36 BY SIMILARITY.
SQ SEQUENCE 37 AA; 3951 MW; EEC16AE67EE3F36A CRC64;
Query Match 69.4%; Score 34; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 KYCCS 6
Db 15 KYCCS 19

Search completed: November 1, 2002, 14:47:51
Job time : 2.41158 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:58 ; Search time 0.848875 seconds
(without alignments)
905.569 Million cell updates/sec

Title: us-09-606-129a-35

Perfect score: 49

Sequence: 1 QKYCCSRK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues.

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	296	2 G02066	biliverdin reducta
2	49	100.0	296	2 S62624	biliverdin reducta
3	38	77.6	151	2 S23215	hypothetical prote
4	38	77.6	703	2 B82148	ATP-dependent heli
5	37	75.5	571	2 T27220	hypothetical prote
6	36	73.5	71	2 T43545	DNA-directed RNA p
7	36	73.5	71	2 C90129	Putative RNA polym
8	36	73.5	150	2 S47354	ribosomal protein
9	36	73.5	263	2 T32330	hypothetical prote
10	36	73.5	761	1 B64506	DNA topoisomerase
11	35	71.4	76	2 E64324	DNA-directed RNA p
12	35	71.4	88	2 T07013	glycine-rich prote
13	35	71.4	127	2 T20707	hypothetical prote
14	35	71.4	202	2 A53399	L6 antigen - mouse
15	35	71.4	243	2 S34794	osmotin - common t
16	35	71.4	385	2 T33038	hypothetical prote
17	35	71.4	709	1 TWBYS5	transcription fact
18	35	71.4	859	2 T38168	hypothetical prote
19	35	71.4	959	2 T23094	hypothetical prote
20	35	71.4	961	2 T23095	hypothetical prote
21	35	71.4	2414	2 A54277	transcription adap
22	35	71.4	2440	2 S39162	transcription coac
23	35	71.4	2441	2 S39161	CREB-binding prote
24	35	71.4	3190	2 T13828	CREB-binding prote
25	34	69.4	73	1 ARRA5G	allergen Ra5G prec
26	34	69.4	129	2 S69304	probable membrane
27	34	69.4	236	2 E97101	protein from bacte
28	34	69.4	418	2 S77640	exopolysaccharide
29	34	69.4	485	2 C86336	hypothetical prote

ALIGNMENTS

RESULT 1

G02066

biliverdin reductase (EC 1.3.1.24) - human

N:Alternate names: biliverdin IX-alpha reductase

C:Species: Homo sapiens (man)

C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998

C:Accession: G02066

R:Komuro, A.; Tobe, T.; Nakano, Y.; Yamaguchi, T.; Tomita, M.

submitted to the EMBL Data Library, August 1995

A:Reference number: H00768

A:Accession: G02066

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-296 <KOM>

A:Cross-references: EMBL:U34877; MID:g1143231; PID:g1143232

C:Keywords: oxidoreductase

Query Match 100.0%; Score 49; DB 2; Length 296;

Best Local Similarity 100.0%; Pred. NO. 0.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QKYCCSRK 8

|||||||

Db 289 QKYCCSRK 296

RESULT 2

S62624

biliverdin reductase (EC 1.3.1.24) - human

N:Alternate names: biliverdin IX-alpha reductase

C:Species: Homo sapiens (man)

C>Date: 28-Oct-1996 #sequence_revision 09-May-1997 #text_change 21-Jul-2000

C:Accession: S62624; S62622; S29736

R:Maines, M.D.; Polevoda, B.V.; Huang, T.J.; McCoubrey Jr., W.K.

Eur. J. Biochem. 235, 372-381, 1996

A:Title: Human biliverdin IX-alpha reductase is a zinc-metalloprotein. Characterizati

A:Reference number: S62622; MUID:96202961

A:Accession: S62624

A:Molecule type: mRNA

A:Residues: 1-296 <MAI>

A:Cross-references: EMBL:X93086; NID:g1246748; PIDN:CAA63635.1; PID:g1246749

A:Accession: S62622

A:Molecule type: protein

A:Residues: 3-24, 'X', 26-27, 'X', 29-36; 48-74; 228-234; 235-248 <MAF>

R:Maines, M.D.; Trakshel, G.M.

Arch. Biochem. Biophys. 300, 320-326, 1993

A:Title: Purification and characterization of human biliverdin reductase.

A:Reference number: S29736; MUID:93143333

A:Accession: S29736

A:Molecule type: protein

A:Residues: 3-24, 'X', 26-27, 'X', 29-36; 48-74; 228-234; 235-248 <MAW>

A>Note: the sequence of peptide 1 from page 323 seems not to belong to this protein

hypothetical prote
hypothetical prote
genome polyprotein
genome polyprotein
hypothetical prote
alpha-2-macroglobu
DNA-directed RNA p
ribosomal protein
probable DNA-direc
hypothetical prote
probable membrane
hypothetical prote
hypothetical prote
hypothetical prote
pathogenesis-relat

C;Genetics:
A;Gene: BVR
C;Keywords: oxidoreductase
F;3-296/Product: biliverdin reductase IX-alpha #status experimental <MAT>

Query Match 100.0%; Score 49; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
|:|||||
Db 289 QKYCCSRK 296

RESULT 3
S23215
hypothetical protein 1 - Photobacterium sp.
C;Species: Photobacterium sp.
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S23215
R;Bartlett, D.H.; Chi, E.; Wright, M.E.
submitted to the EMBL Data Library, June 1992
A;Description: Nucleotide sequence of the OMPH gene and construction of an OMPH mutant
A;Reference number: S23213
A;Accession: S23215
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-151 <BAR>
A;Cross-references: EMBL:X67094; NID:g45887; PIDN:CAA47468.1; PID:g45890

Query Match 77.6%; Score 38; DB 2; Length 151;
Best Local Similarity 62.5%; Pred. No. 28;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
|:|||||
Db 119 QKYCCNKH 126

RESULT 4
B82148
ATP-dependent helicase, Dmg family VC1855 [imported] - Vibrio cholerae (strain N16961)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: B82148
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Esmolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H.
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: B82148
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-703 <HEI>
A;Cross-references: GB:AE003852; NID:g9656382; PIDN:AAF95003.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC1855
A;Map position: 1

Query Match 77.6%; Score 38; DB 2; Length 703;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
|:|||||
Db 131 QKYCCSRK 138

RESULT 5
T27220
hypothetical protein Y57G11C.7 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27220
R;McMurray, A.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z20330
A;Accession: T27220
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-571 <WIL>
A;Cross-references: EMBL:Z99281; PIDN:CAB16509.1; GSPDB:GN000022; CESP:Y57G11C.7
A;Experimental source: clone Y57G11C
C;Genetics:
A;Gene: CESP:Y57G11C.7
A;Map position: 4
A;Introns: 72/2; 155/3; 214/1; 249/3; 279/3

Query Match 75.5%; Score 37; DB 2; Length 571;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
|:|||||
Db 192 RKYCCPRR 199

RESULT 6
T43545
DNA-directed RNA polymerase (EC 2.7.7.6) II chain Rpb10 - fission yeast (Schizosaccha
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 03-Nov-2000
C;Accession: T43545; T38031; T52005
R;Shpakovski, G.V.; Lebedenko, E.N.; Proshkin, S.A.
submitted to the EMBL Data Library, October 1997
A;Reference number: Z22564
A;Accession: T43545
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-71 <SHP>
A;Cross-references: EMBL:AF027818; PIDN:AAC16895.1
A;Experimental source: strain 9721(-)
R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z21763
A;Accession: T38031
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-71 <MUR>
A;Cross-references: EMBL:Z98598; PIDN:CAB11246.1; GSPDB:GN000066; SPDB:SPAC1B3.12C
A;Experimental source: strain 972h-; cosmid c1B3
R;Sakurai, H.; Ishihama, A.
Gene 196, 165-174, 1997
A;Title: Gene organization and protein sequence of the small subunits of Schizosaccha
A;Reference number: Z25897; MUID:97464063
A;Accession: T52005
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-71 <SAK>
A;Cross-references: EMBL:D89596; PIDN:BAA22805.1
A;Experimental source: strain JY741
C;Genetics:
A;Gene: rpb10; SPAC1B3.12C
A;Map position: 1
A;Introns: 12/3; 64/1
A;Note: rpb10
C;Superfamily: DNA-directed RNA polymerase II chain RPB10
C;Keywords: nucleotidyltransferase; transcription

Query Match 73.5%; Score 36; DB 2; Length 71;
Best Local Similarity 71.4%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSR 7

Db 41 QRYCCRR 47
|:|||||

RESULT 7
C90129

Putative RNA polymerase II subunit Rpb10 [Imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: C90129
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reijnders, L.J.; et al. 2001
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; MUID:11323671
A:Accession: C90129
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-71 <DOU>
A:Cross-references: GB:AF083031; NID:gl3794349; PIDN:AAK39726.1; GSPDB:GN00152
C:Genetics:
A:Gene: rpb10
A:Map position: 3
A:Genome: nucleomorph
C:Keywords: nucleomorph

Query Match 73.5%; Score 36; DB 2; Length 71;
Best Local Similarity 71.4%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKYCCSR 7
|:|||||

Db 41 QRYCCRR 47

RESULT 8
S47354

ribosomal protein L18a homolog - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 24-Oct-1997 #text_change 05-Nov-1999
C:Accession: S47354
R:Zenz, K.I.
submitted to the EMBL Data Library, August 1994
A:Reference number: S47353
A:Accession: S47354
A:Molecule type: mRNA
A:Residues: 1-150 <ZEN>
A:Cross-references: EMBL:X80821; NID:g527579; PIDN:CAA56787.1; PID:g527580

Query Match 73.5%; Score 36; DB 2; Length 150;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QKYCCSRK 8
|:|||||

Db 135 QKYCCLLK 142

RESULT 9
T32530

hypothetical protein CD4.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32530
R:Du, Z.; Scheet, P.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid CD4.
A:Reference number: Z21185
A:Accession: T32530
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-263 <DUZ>
A:Cross-references: EMBL:AF036694; PIDN:AAB88345.1; GSPDB:GN00022; CESP:CD4.7

A:Experimental source: strain Bristol N2; clone CD4
C:Genetics:
A:Gene: CESP:CD4.7
A:Map position: 4
A:Introns: 121/3; 156/1; 243/2

Query Match 73.5%; Score 36; DB 2; Length 263;
Best Local Similarity 83.3%; Pred. No. 81;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KYCCSR 7
|:|||||

Db 234 KYCCTR 239

RESULT 10
B64506

DNA topoisomerase (EC 5.99.1.2) - Methanococcus jannaschii
N:Alternate names: topoisomerase I
C:Species: Methanococcus jannaschii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: B64506
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: B64506
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-761 <BUL>
A:Cross-references: GB:U67605; GB:L77117; NID:gl592227; PIDN:AAB99673.1; PID:gl592234
C:Genetics:
A:Map position: FOR1633497-1635782
A:Start codon: GTG
C:Superfamily: DNA topoisomerase I topA
C:Keywords: isomerase

Query Match 73.5%; Score 36; DB 1; Length 761;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KYCCSRK 8
|:|||||

Db 125 KYCCGRE 131

RESULT 11
E64324

DNA-directed RNA polymerase (EC 2.7.7.6) subunit N - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 23-Apr-1999
C:Accession: E64324
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: E64324
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-76 <BUL>
A:Cross-references: GB:U67475; GB:L77117; NID:gl590930; PID:gl590941; TIGR:MJ0196; PI
C:Genetics:
A:Map position: FOR190573-190803
A:Start codon: GTG
C:Superfamily: DNA-directed RNA polymerase II chain RPB10
C:Keywords: nucleotidyltransferase; transcription

Query Match 71.4%; Score 35; DB 2; Length 76;
Best Local Similarity 71.4%; Pred. No. 53;
Matches 5; Conservative 1; Mismatches 0; Gaps 0;

QY 1 QKYCCSR 7
:|||||
Db 44 KKYCCRR 50

RESULT 12

T07013
glycine-rich protein precursor - tomato
N;Alternate names: cell wall protein
C;Species: Lycopersicon esculentum (tomato)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-May-2000
C;Accession: T07013
R;Domingo, C.; Gomez, M.D.; Canas, L.; Hernandez-Yago, J.; Conejero, V.; Vera, P.
Plant Cell 6, 1035-1047, 1994
A;Title: A novel extracellular matrix protein from tomato plant associated with lignification
A;Reference number: Z15861; MUID:95003699
A;Accession: T07013
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-88 <DOM>
A;Cross-references: EMBL:X77373; NID:g541681; PIDN:CAA54561.1; PID:g541682
A;Experimental source: cultivar Rutgers; expanded laeves
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-88/Product: glycine-rich protein #status predicted <MAT>

Query Match 71.4%; Score 35; DB 2; Length 88;
Best Local Similarity 83.3%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCS 6
:|||||
Db 72 KKYCCS 77

RESULT 13

T20707
hypothetical protein F10D11.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20707
R;Dobson, R.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19312
A;Accession: T20707
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-127 <WIL>
A;Cross-references: EMBL:Z81057; PIDN:CAB02915.1; GSPDB:GN000019; CESP:F10D11.3
A;Experimental source: clone F10D11
C;Genetics:
A;Gene: CESP:F10D11.3
A;Map position: 1
A;Introns: 22/3; 52/3; 98/3

Query Match 71.4%; Score 35; DB 2; Length 127;
Best Local Similarity 83.3%; Pred. No. 74;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YCCSRK 8
:|||||
Db 118 FCCSRK 123

RESULT 14

A53399
L6 antigen - mouse
C;Species: Mus musculus (house mouse)
C;Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 05-Nov-1999
C;Accession: A53399

R;Marken, J.S.; Bajorath, J.; Edwards, C.P.; Farr, A.G.; Schieven, G.L.; Hellstroem, J. Biol. Chem. 269, 7397-7401, 1994
A;Title: Membrane topology of the L6 antigen and identification of the protein epitop
A;Reference number: A53399; MUID:94171760
A;Accession: A53399
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-202 <MAR>
A;Cross-references: GB:I15429; NID:g468909; PIDN:AAA17873.1; PID:g476343

Query Match 71.4%; Score 35; DB 2; Length 202;
Best Local Similarity 83.3%; Pred. No. 99;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YCCSRK 8
:|||||
Db 192 YCCSRQ 197

RESULT 15

S34794
osmotin - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C;Accession: S34794; S37294
R;Singh, N.K.; Nelson, D.E.; Kuhn, D.; Hasegawa, P.M.; Bressan, R.A.
Plant Physiol. 90, 1096-1101, 1989
A;Title: Molecular cloning of osmotin and regulation of its expression by ABA and ada
A;Reference number: S34794
A;Accession: S34794
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-243 <SIN>
A;Cross-references: EMBL:M29279
R;Singh, N.K.; Nelson, D.E.; Kuhn, D.; Hasegawa, P.M.; Bressan, R.A.
submitted to the EMBL Data Library, February 1990
A;Reference number: S37294
A;Accession: S37294
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-128,'1',130-243 <SI2>
A;Cross-references: EMBL:M29279; NID:gl70278; PIDN:AAA34089.1; PID:gl70279
C;Superfamily: thaumatin I

Query Match 71.4%; Score 35; DB 2; Length 243;
Best Local Similarity 50.0%; Pred. No. 11e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
:|||||
Db 171 QKYCCTQR 178

Search completed: November 1, 2002, 14:51:03
Job time : 3.01554 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:44:48 ; Search time 0.66881 Seconds
(without alignments)
292.168 Million cell updates/s

Title: US-09-606-129A-35
Perfect score: 49
Sequence: 1 QKYCCSRK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database :
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	COMPARISON			Description
			DB	ID		
1	36	73.5	109	2	US-08-527-044-2	Sequence 2, Appli
2	36	73.5	109	3	US-09-013-780-2	Sequence 2, Appli
3	35	71.4	202	2	US-08-855-261A-4	Sequence 4, Appli
4	35	71.4	202	3	US-08-839-711-4	Sequence 4, Appli
5	35	71.4	202	4	US-09-227-224-4	Sequence 4, Appli
6	35	71.4	2414	1	US-08-227-536-2	Sequence 2, Appli
7	35	71.4	2414	5	PCT-US95-04682-2	Sequence 2, Appli
8	35	71.4	2441	1	US-08-194-468-2	Sequence 2, Appli
9	35	71.4	2441	3	US-08-961-739-2	Sequence 2, Appli
10	34	69.4	117	3	US-08-965-903B-6	Sequence 2, Appli
11	34	69.4	139	3	US-08-965-903B-20	Sequence 20, Appli
12	34	69.4	174	4	US-09-383-586-12	Sequence 12, Appli
13	34	69.4	1693	3	US-08-840-316-1	Sequence 1, Appli
14	34	69.4	1693	3	US-08-478-507-7	Sequence 7, Appli
15	34	69.4	1693	4	US-08-809-523-1	Sequence 1, Appli
16	34	69.4	1693	4	US-09-128-275A-7	Sequence 7, Appli
17	34	69.4	1693	4	US-08-471-971-1	Sequence 1, Appli
18	34	69.4	1693	5	PCT-US93-08849A-1	Sequence 1, Appli
19	34	69.4	1693	5	PCT-US93-08849-1	Sequence 1, Appli
20	33	67.3	24	4	US-08-857-076-85	Sequence 85, Appli
21	33	67.3	205	4	US-08-857-076-113	Sequence 113, Appli
22	33	67.3	226	1	US-07-828-798C-7	Sequence 7, Appli
23	33	67.3	226	2	US-08-315-868A-7	Sequence 7, Appli
24	33	67.3	226	3	US-08-495-8195-7	Sequence 7, Appli
25	33	67.3	246	1	US-07-828-798C-6	Sequence 6, Appli
26	33	67.3	246	2	US-08-315-868A-6	Sequence 6, Appli
27	33	67.3	246	3	US-08-495-819B-6	Sequence 6, Appli

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 73.5%; Score 36; DB 2; Length 109;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5: Conservative 1; Mismatches 0; Indels

QY	2 KYCCSR	7
	:	
Db	22 KYCCSK	27

RESULT 2

```

US-09-013-780-2
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0298 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; ADDRESS: Akzo No. 6001363el N.V
; STREET: 1300 PICCARD DRIVE #206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850-4373
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,780
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/527,044
; FILING DATE: 12-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GORMLEY, MARY E.
; REGISTRATION NUMBER: 34,409
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-013-780-2
;
Query Match 73.5%; Score 36; DB 3; Length 109;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KYCCSR 7
Db 22 KYCCSK 27

RESULT 3
US-08-855-261A-4
; Sequence 4, Application US/08855261A
; Patent No. 5922566
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW TUMOR-ASSOCIATED ANTIGEN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,711
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0270 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 202 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; US-08-855-261A-4

US-09-013-780-2
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0298 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; ADDRESS: Akzo No. 6001363el N.V
; STREET: 1300 PICCARD DRIVE #206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850-4373
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,780
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/527,044
; FILING DATE: 12-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GORMLEY, MARY E.
; REGISTRATION NUMBER: 34,409
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-013-780-2
;
Query Match 71.4%; Score 35; DB 2; Length 202;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YCCSRK 8
Db 192 YCCSRQ 197

RESULT 4
US-08-839-711-4
; Sequence 4, Application US/08839711
; Patent No. 6033870
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,711
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0270 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 202 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; US-08-839-711-4

```

; LIBRARY: GenBank
; CLONE: 476343
; US-08-839-711-4

Query Match 71.4%; Score 35; DB 3; Length 202;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YCCSRK 8

|||||

Db 192 YCCSRQ 197

RESULT 5

US-09-227-224-4

; Sequence 4, Application US/09227224

; Patent No. 6350581

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: NEW TUMOR-ASSOCIATED ANTIGEN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/227,224

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/855,261

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0298 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 202 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 476343

US-09-227-224-4

Query Match

Best Local Similarity 71.4%; Score 35; DB 4; Length 202;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 YCCSRK 8

|||||

Db 192 YCCSRQ 197

RESULT 6

US-08-227-536-2

; Sequence 2, Application US/08227536

; Patent No. 5658784

; GENERAL INFORMATION:
; APPLICANT: Eckner, Richard
; APPLICANT: Ewen, Mark
; APPLICANT: Livingston, David
; TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
; TITLE OF INVENTION: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,536
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Ph.D., Kathleen A.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: DFCI-308XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2414 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-227-536-2

Query Match 71.4%; Score 35; DB 1; Length 2414;

Best Local Similarity 83.3%; Pred. No. 8e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 YCCSRK 8

|||||

Db 1162 YCCGRK 1167

RESULT 7

PCT-US95-04682-2

; Sequence 2, Application PC/TUS9504682

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION

; TITLE OF INVENTION: FACTOR P300 AND USES OF P300

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

; STREET: Ten Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/04682

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/227,536

```
; FILING DATE: 14-April-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Holliday C. Heine, Ph.D.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: DFCI-308Xq999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2414 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-04682-2

Query Match 71.4%; Score 35; DB 5; Length 2414;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 YCCSRK 8
   ||| ||
Db 1162 YCCGRK 1167

RESULT 8
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; TITLE OF INVENTION: RESPONSIVE GENES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-194-468-2

Query Match 71.4%; Score 35; DB 1; Length 2441;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 YCCSRK 8
   ||| ||
```

```
Db 1199 YCCGRK 1204

RESULT 9
US-08-961-739-2
; Sequence 2, Application US/08961739A
; Patent No. 6063583
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
; FILE REFERENCE: SALK1650-1
; CURRENT APPLICATION NUMBER: US/08/961,739A
; CURRENT FILING DATE: 1997-10-31
; EARLIER APPLICATION NUMBER: US 194,468
; EARLIER FILING DATE: 1994-02-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: Mus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2441)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-08-961-739-2

Query Match 71.4%; Score 35; DB 3; Length 2441;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 YCCSRK 8
   ||| ||
Db 1199 YCCGRK 1204

RESULT 10
US-08-965-903B-6
; Sequence 6, Application US/08965903B
; Patent No. 6060275
; GENERAL INFORMATION:
; APPLICANT: Hacothen, Nir
; APPLICANT: Krasnow, Mark A.
; TITLE OF INVENTION: SPROUTY PROTEIN AND CODING
; TITLE OF INVENTION: SEQUENCE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,903B
; FILING DATE: 07-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030232
; FILING DATE: 07-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Petithory, Joanne R
; REGISTRATION NUMBER: 42,995
; REFERENCE/DOCKET NUMBER: 8600-0177.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; TELEX:
```

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: h-spryl cysteine-rich region
US-08-965-903B-6

Query Match 69.4%; Score 34; DB 3; Length 117;
Best Local Similarity 71.4%; Pred. No. 91;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKYCCSR 7
| :||||
Db 70 QSHCCSR 76

RESULT 11

US-08-965-903B-20
Sequence 20, Application US/08965903B
Patent No. 6060275

GENERAL INFORMATION:
APPLICANT: Hacothen, Nir
APPLICANT: Krasnow, Mark A.
TITLE OF INVENTION: SPROUTY PROTEIN AND CODING
TITLE OF INVENTION: SEQUENCE
NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA

COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/965,903B
FILING DATE: 07-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/030232
FILING DATE: 07-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Petithory, Joanne R
REGISTRATION NUMBER: 42,995
REFERENCE/DOCKET NUMBER: 8600-0177,30
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
TELEX:

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: h-spryl
US-08-965-903B-20

Query Match 69.4%; Score 34; DB 3; Length 139;
Best Local Similarity 71.4%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKYCCSR 7
| :||||
Db 70 QSHCCSR 76

RESULT 12

US-09-383-586-12
Sequence 12, Application US/09383586
Patent No. 6242419

GENERAL INFORMATION:
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Abernethy, Nevin
APPLICANT: Onrust, Rene
APPLICANT: Kumble, Anand
APPLICANT: Murison, Greg
TITLE OF INVENTION: Compounds isolated from stromal cells
TITLE OF INVENTION: and methods for their use
FILE REFERENCE: 11000.1037c1
CURRENT APPLICATION NUMBER: US/09/383,586
CURRENT FILING DATE: 1999-08-26
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 12

LENGTH: 174
TYPE: PRT
ORGANISM: Mouse
US-09-383-586-12

Query Match 69.4%; Score 34; DB 4; Length 174;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 YCCSRK 8
||| ||
Db 116 YCCRRK 121

RESULT 13

US-08-840-316-1
Sequence 1, Application US/08840316
Patent No. 6054567

GENERAL INFORMATION:

APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
APPLICANT: Tsarev, Sergei. A., and Robinson, Robin A.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,316
FILING DATE: 11-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Richard W. Bork
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4255

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 AMINO ACID RESIDUES
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
US-08-840-316-1

Query Match 69.4%; Score 34; DB 3; Length 1693;
Best Local Similarity 71.4%; Pred. No. 8.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 QKYCCSR 7
| :||||
Db 333 QAFCCSR 339

RESULT 14

US-08-478-507-7
Sequence 7, Application US/08478507
Patent No. 6120988
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarbough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk E
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-507-7

Query Match 69.4%; Score 34; DB 3; Length 1693;
Best Local Similarity 71.4%; Pred. No. 8.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 QKYCCSR 7
| :||||
Db 333 QAFCCSR 339

RESULT 15

US-08-809-523-1
Sequence 1, Application US/08809523
Patent No. 6207416
GENERAL INFORMATION:
APPLICANT: Tsarev, Sergei A., Emerson,
APPLICANT: Suzanne U., Purcell, Robert H.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/809,523
FILING DATE: 28-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13102
FILING DATE: 03-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US08/316,765
FILING DATE: 03-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,263
FILING DATE: 18-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Richard W. Bork
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-40320S4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 AMINO ACID RESIDUES
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
US-08-809-523-1

Query Match 69.4%; Score 34; DB 4; Length 1693;
Best Local Similarity 71.4%; Pred. No. 8.3e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QKYCCSR 7

| :||||

Db 333 QAFCCSR 339

Search completed: November 1, 2002, 14:52:00
Job time : 2.6681 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:40:53 ; Search time 1.8135 Seconds
(without alignments)
489.985 Million cell updates/sec

Title: US-09-606-129A-35
Perfect score: 49
Sequence: 1 QKYCCSRK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	79.6	74	21	AA844795
2	39	79.6	119	21	AA844820
3	39	79.6	119	21	AA844821
4	39	79.6	124	22	AA674903
5	36	73.5	58	22	ABB69460
6	36	73.5	109	17	AA844086
7	35	73.5	505	22	ABC27093
8	35	71.4	124	22	AA899956
9	35	71.4	189	20	AAV34717
10	35	71.4	246	19	AAW69751
11	35	71.4	246	20	AAW94510

12	35	71.4	316	22	ABG09140	Novel human diaqno
13	35	71.4	345	21	AA825483	Eucalyptus grandis
14	35	71.4	2414	16	AA884882	Transcription fact
15	35	71.4	2414	19	AAW40057	Cellular transcript
16	35	71.4	2441	16	AA879054	CREB binding prote
17	35	71.4	2441	19	AAW40058	Cellular transcript
18	35	71.4	2441	21	AAV94252	Mouse nuclear CREB
19	35	71.4	2441	22	ABB44555	Mouse wound healin
20	35	71.4	2442	22	ABB44556	Human wound healin
21	35	71.4	3190	22	AA884634	Amino acid sequenc
22	35	71.4	3275	22	ABB70437	Drosophila melanog
23	34	69.4	57	22	AAW86342	Human immune/haema
24	34	69.4	60	19	AAW74947	Human secreted pro
25	34	69.4	61	19	AAW74790	Human secreted pro
26	34	69.4	90	22	ABB15777	Human nervous syst
27	34	69.4	100	22	AAU56930	Protonibacterium
28	34	69.4	117	19	AAW48794	Homo sapiens sprou
29	34	69.4	136	21	AAV58843	Maize resistance 9
30	34	69.4	139	21	AAW41184	Human ORFX ORF948
31	34	69.4	140	19	AAW48793	Homo sapiens sprou
32	34	69.4	156	21	AAV87596	Human SPROUTY-1 pr
33	34	69.4	167	21	AAV87595	Human SPROUTY-1 pr
34	34	69.4	172	21	AAV58851	Sorghum resistance
35	34	69.4	174	21	AAW19104	Polypeptide isolat
36	34	69.4	288	21	AAV87597	Human SPROUTY prot
37	34	69.4	294	21	AAV87592	Human SPROUTY-1 pr
38	34	69.4	295	21	AAV87593	Human SPROUTY-1 pr
39	34	69.4	319	21	AAV87591	Human SPROUTY-1 pr
40	34	69.4	319	22	AAW38688	Human polypeptide
41	34	69.4	319	22	AAW64377	Amino acid sequenc
42	34	69.4	326	22	AAW40474	Human polypeptide
43	34	69.4	338	21	AAV87594	Human SPROUTY-1 pr
44	34	69.4	423	21	AAW23211	Arabidopsis thalia
45	34	69.4	485	21	AAW23210	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAB44795
ID AAB44795 standard; Protein; 74 AA.
XX
AC AAB44795;
XX
DT 12-FEB-2001 (first entry)
XX
Human secreted protein sequence encoded by gene 34 SEQ ID NO:94.
XX
Human: secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX
antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX
cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
XX
fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
XX
hyperproliferative disorder; cardiovascular disorder; angiogenesis;
XX
cerebrovascular disorder; nervous system disorder; ocular disorder;
XX
wound healing; skin aging; food additive; preservative.
XX
OS Homo sapiens.
XX
PN WO200058336-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07726.
XX
PR 26-MAR-1999; 99US-0126597.
XX
PR 07-JAN-2000; 2000US-0174877.
XX
(HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-602355/57.

DR N-PSDB; AAC79832.
XX Nucleic acid encoding human secreted proteins, used to treat, prevent,
PT ameliorate or diagnose medical conditions such as cancer, and
PT autoimmune diseases -
XX
PS Claim 11; Page 368-369; 391pp; English.
XX
XX The polynucleotide sequences given in AAC79799 to AAC79848 encode the
CC human secreted proteins given in AAB44762 to AAB44811. AAB44812 to
CC AAB44829 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are used in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities are:
CC immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC neuroprotective; antibacterial; virucide; fungicide; and
CC ophthalmological. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. AAC79790 to AAC79798 and
CC AAB44761 represent sequences used in the exemplification of the present
XX invention.
XX
SQ Sequence 74 AA;

Query Match 79.6%; Score 39; DB 21; Length 74;
Best Local Similarity 75.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
:|||||
Db 57 EKYCCSPK 64

RESULT 2
AAB44820
ID AAB44820 standard; Protein; 119 AA.
XX
XX AAB44820;
XX
XX 12-FEB-2001 (first entry)
XX
XX Gene 34 human secreted protein homologous amino acid sequence #119.
DE
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW cerebrovascular disorder; nervous system disorder; ocular disorder;
KW wound healing; skin aging; food additive; preservative.
XX
XX Homo sapiens.
OS
XX WO200058336-A1.
PN
XX 05-OCT-2000.
PD
XX
XX 23-MAR-2000; 2000WO-0507726.
PF
XX 26-MAR-1999; 99US-0126597.
XX
PR 07-JAN-2000; 2000US-0174877.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-602355/57.
XX
XX Nucleic acid encoding human secreted proteins, used to treat, prevent,
PT ameliorate or diagnose medical conditions such as cancer, and
PT autoimmune diseases -
XX
XX Disclosure; Page 379-380; 391pp; English.
XX
XX The polynucleotide sequences given in AAC79799 to AAC79848 encode the
CC human secreted proteins given in AAB44762 to AAB44811. AAB44812 to
CC AAB44829 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are used in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities are:
CC immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC neuroprotective; antibacterial; virucide; fungicide; and
CC ophthalmological. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. AAC79790 to AAC79798 and
CC AAB44761 represent sequences used in the exemplification of the present
XX invention.
XX
SQ Sequence 119 AA;

Query Match 79.6%; Score 39; DB 21; Length 119;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QKYCCSRK 8
:|||||
Db 103 EKYCCSPK 110

RESULT 3
AAB44821
ID AAB44821 standard; Protein; 119 AA.
XX
XX AAB44821;
XX
XX 12-FEB-2001 (first entry)
XX
XX Human secreted protein sequence encoded by gene 34 SEQ ID NO:120.
DE
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW cerebrovascular disorder; nervous system disorder; ocular disorder;
KW wound healing; skin aging; food additive; preservative.
XX
XX Homo sapiens.
OS
XX WO200058336-A1.
PN
XX 05-OCT-2000.
PD

```
XX 23-MAR-2000; 2000WO-US07726.
XX PF
XX PR
XX PR 26-MAR-1999; 99US-0126597.
XX PR 07-JAN-2000; 2000US-0174877.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX PI WPI; 2000-602355/57.
XX DR
XX PT Nucleic acid encoding human secreted proteins, used to treat, prevent,
XX PT ameliorate or diagnose medical conditions such as cancer, and
XX PT autoimmune diseases -
XX PS
XX PS Disclosure; Page 380; 391pp; English.
XX CC The polynucleotide sequences given in AAC79799 to AAC79848 encode the
XX CC human secreted proteins given in AAB44762 to AAB44811. AAB44812 to
XX CC AAB44829 represent human secreted polypeptide sequences and proteins
XX CC homologous to them, which are used in the exemplification of the present
XX CC invention. Human secreted proteins have activities based on the tissues
XX CC and cells the genes are expressed in. Examples of activities are:
XX CC immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
XX CC cytostatic; cardiac; antibacterial; cerebroprotective; neurotropic;
XX CC neuroprotective; antiviral; virucide; fungicide; and
XX CC ophthalmological. The polynucleotides and polypeptides can be used to
XX CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX CC in diagnosing a pathological condition or susceptibility to a
XX CC pathological condition. Disorders which are diagnosed or treated include
XX CC autoimmune diseases, hyperproliferative disorders, cardiovascular
XX CC disorders, cerebrovascular disorders, angiogenesis, nervous system
XX CC disorders, infections caused by bacteria, viruses and fungi and ocular
XX CC disorders. The polypeptides can also be used to aid wound healing and
XX CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
XX CC maintain organs before transplantation, for supporting cell culture of
XX CC primary tissues, to regenerate tissues and in chemotaxis. The
XX CC polypeptides can also be used as a food additive or preservative to
XX CC increase or decrease storage capabilities. AAC79790 to AAC79798 and
XX CC AAB44761 represent sequences used in the exemplification of the present
XX CC invention.
XX SQ Sequence 119 AA;
Query Match 79.6%; Score 39; DB 21; Length 119;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 QKYCCSRK 8
Db 103 EKYCCSPK 110
:|||||
RESULT 4
AAG74903
ID AAG74903 standard; Protein; 124 AA.
XX AC AAG74903;
XX XX
XX DT 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen protein SEQ ID NO:5667.
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma.
XX OS Homo sapiens.
XX PN WO200122920-A2.
XX PR 05-APR-2001.
```

```
XX 28-SEP-2000; 2000WO-US26524.
XX PF
XX PR
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX PI WPI; 2001-235357/24.
XX DR N-PSDB; AAB34308.
XX XX
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX PS
XX PS Claim 11; Page 7210-7211; 9803pp; English.
XX CC AAB32943 to AAB37195 and AAG73514 to AAG77788 represent human colon
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
XX CC the proteins are collectively known as colon cancer antigens. The colon
XX CC cancer antigens have cytostatic activity and can be used in gene
XX CC therapy and vaccine production. N and P may be used in the prevention,
XX CC diagnosis and treatment of diseases associated with inappropriate P
XX CC expression. For example, N and P may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of P by expressing
XX CC inactive proteins or to supplement the patients own production of P.
XX CC Additionally, N may be used to produce the colon cancer-associated Ps,
XX CC by inserting the nucleic acids into a host cell and culturing the cell
XX CC to express the proteins. N and P can be used in the prevention, diagnosis
XX CC and treatment of colorectal carcinomas and cancers. AAB37196 to AAB37204
XX CC and AAB77789 represent sequences used in the exemplification of the
XX CC present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX CC missing at time of publication, meaning no sequences are present for
XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX SQ Sequence 124 AA;
Query Match 79.6%; Score 39; DB 22; Length 124;
Best Local Similarity 75.0%; Pred. No. 58;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 QKYCCSRK 8
Db 108 EKYCCSPK 115
:|||||
RESULT 5
ABB69460
ID ABB69460 standard; Protein; 58 AA.
XX AC ABB69460;
XX XX
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 35172.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PR 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
```

PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI: 2001-656860/75.
 DR N-PSDB; ABL13563.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions .
 XX
 XX Disclosure; SEQ ID NO 35172; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 58 AA;
 SQ
 Query Match 73.5%; Score 36; DB 22; Length 58;
 Best Local Similarity 71.4%; Pred. No. 90;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 KYCCSRK 8
 Db 18 KYCCKAEK 24
 IIII: I
 RESULT 6
 AAR84086
 ID AAR84086 standard; Protein; 109 AA.
 XX
 AC AAR84086;
 XX
 DT 28-NOV-1996 (first entry)
 XX
 DE T-lymphocyte stimulatory protein.
 XX
 KW E. maxima; T-lymphocyte stimulatory protein; Eimeria; protozoan;
 KW coccidiosis; chicken; vaccine; poultry; probe.
 XX
 OS Eimeria maxima.
 XX
 PN AU9531720-A.
 XX
 XX 28-MAR-1996.
 PD
 XX 15-SEP-1995; 95AU-0031720.
 PF
 XX 16-SEP-1994; 94EP-0202676.
 PR
 XX (ALKU) AKZO NOBEL NV.
 PA
 XX Bumstead JM, Dunn PJJ, Tomley FM, Vermeulen AN;
 PI
 XX WPI: 1996-210114/22.
 DR N-PSDB; AAT14351.
 XX
 XX DNA encoding Eimeria T-lymphocyte stimulatory protein - used in
 PT vaccines to protect poultry against coccidiosis, and to develop
 PT prods. for diagnosis of Eimeria infection
 XX
 XX Claim 1; Page 46; 59pp; English.
 PS
 XX This sequence represents E. maxima T-lymphocyte stimulatory protein.
 CC

CC Eimeria protozoans are the cause of coccidiosis in chickens. The DNA
 CC encoding this sequence may be attached to a suitable promoter and used
 CC in a recombinant vector in the production of a vaccine for the protection
 CC of poultry against coccidiosis. Fragments of this sequence may also
 CC be used as probes to detect Eimeria-related nucleic acid in tissue.
 CC due to poor print quality in the specification, this sequence is a
 CC "best guess" based on the corresponding DNA sequence.
 XX
 XX Sequence 109 AA;
 SQ
 Query Match 73.5%; Score 36; DB 17; Length 109;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 KYCCSR 7
 Db 22 KYCCSK 27
 IIII: I
 RESULT 7
 ABG27093
 ID ABG27093 standard; Protein; 505 AA.
 XX
 AC ABG27093;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #27084.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WC200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-0508631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR
 XX 23-AUG-2000; 2000US-0649167.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS91280.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 20; SEQ ID No 57452; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_published_pct_sequences.

XX SQ Sequence 505 AA;

Query Match 73.5%; Score 36; DB 22; Length 505;
 Best Local Similarity 71.4%; Pred. No. 5e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 2; Mismatches 0;

Oy 2 KYCCSRK 8
 |||||:
 Db 322 KYCCSQ 328

RESULT 8
 AAB99956
 ID AAB99956 standard; Protein: 124 AA.

XX AC AAB99956;

XX DT 12-OCT-2001 (first entry)

XX DE Human myelin P0 protein 14 SEQ ID NO:2.

XX KW Human; myelin P0 protein 14; cytostatic; immunomodulatory; haemostatic;
 KW anti-HIV; antiinflammatory; diagnosis; malignant neoplasm; haemopathy;
 KW HIV infection; immunological disease; inflammation.

XX OS Homo sapiens.

XX PN WO200149739-A1.

XX PD 12-JUL-2001.

XX PF 25-DEC-2000; 2000WO-CN00668.

XX PR 29-DEC-1999; 99CN-0127238.

XX PA (UYFU-) UNIV FUDAN.

XX PA (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.

XX PI Mao Y, Xie Y;

XX DR WPI; 2001-432873/46.

XX DR N-PSDB; AAH44460.

XX PT Myelin P0 protein 14 and encoded polynucleotide, applicable in
 PT diagnosis and treatment of malignant neoplasm, haemopathy, HIV
 PT infection, immunological diseases and various inflammation -

XX PS Claim 1; Page 29; 37pp; Chinese.

XX CC The present sequence represents human myelin P0 protein 14: (I). (I) has
 CC cytosolic, immunomodulatory, haemostatic, anti-HIV and antiinflammatory
 CC activities. (I) and the polynucleotide encoding it are applicable in
 CC the diagnosis and treatment of malignant neoplasm, haemopathy, HIV
 CC infection, immunological diseases and various inflammations.

XX SQ Sequence 124 AA;

Query Match 71.4%; Score 35; DB 22; Length 124;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 QKYCC 5

Db 108 QKYCC 112

RESULT 9

AAAY34717
 ID AAY34717 standard; Protein: 189 AA.

XX AC AAY34717;

XX DT 13-SEP-1999 (first entry)

XX DE Chlamydia pneumoniae transmembrane protein sequence.

XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.

XX OS Chlamydia pneumoniae.

XX PN WO9927105-A2.

XX PD 03-JUN-1999.

XX PF 20-NOV-1998; 98WO-IB01890.

XX PR 04-NOV-1998; 98US-0107078.

XX PR 21-NOV-1997; 97FR-0014673.

XX PA (GEST) GENSET.

XX PI Griffais R;

XX DR WPI; 1999-357842/30.

XX PT Genome sequence of Chlamydia pneumoniae

XX PS Page 714-715; Disclosure: 1912pp; English.

XX CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX SQ Sequence 189 AA;

Query Match 71.4%; Score 35; DB 20; Length 189;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 QKYCC 5

Db 17 QKYCC 21

RESULT 10
 AAW69751
 ID AAW69751 standard; Protein: 246 AA.

XX AC AAW69751;

XX DT 02-NOV-1998 (first entry)

XX DE Nicotiana tabacum osmotin protein.

XX KW Nicotiana tabacum; tobacco plant; osmotin gene; promoter; inhibition;
 KW fungal pathogen; insect pathogen; nematode pathogen; viral pathogen.
 XX OS Nicotiana tabacum.

XX FH Key Location/Qualifiers

FT Misc-difference 181
 XX /note= "encoded by GGA"
 XX
 PN US5801028-A.
 XX
 PD 01-SEP-1998.
 XX
 PF 07-JUN-1995; 95US-0482037.
 XX
 PP 12-JAN-1994; 94US-0180428.
 PR 20-MAY-1993; 93US-0065147.
 PR 07-JUN-1995; 95US-0482037.
 XX
 PA (PURD) PURDUE RES FOUND.
 XX
 XX Bressan R, Hasegawa PM;
 PI WPI; 1998-494773/42.
 DR N-PSDB; AAV52754.
 XX
 XX Inhibition of pathogens in plants by recombinant expression of
 PT pathogen inhibiting proteins - uses nucleic acid constructs
 PT containing the pathogen inhibiting proteins under control of osmotin
 PT promoter, which is inducible by specific signals
 XX
 XX Claim 1; Column 29-34; 26pp; English.
 XX
 CC A method has been developed of inhibiting a pathogen in a plant. The
 CC method comprises: (a) providing or constructing a vector comprising an
 CC osmotin promoter and foreign DNA sequence encoding a pathogen-inhibiting
 CC protein, where the promoter is operably linked to the foreign DNA
 CC sequence and includes: (i) nucleotide sequence 5' of the osmotin coding
 CC sequence extending to position -248 bases from the start of the osmotin
 CC protein coding sequence, the coding sequence defined as starting at bp
 CC 2034 of the 3033 bp osmotin sequence (see AAV52754), (ii) a nucleotide
 CC sequence which hybridises to (i) and promotes expression of an operably
 CC linked coding sequence under conditions of desiccation; and (b)
 CC introducing the vector into the plant to create a transformed plant,
 CC where expression of the pathogen-inhibiting protein in the transformed
 CC plant is regulated by the osmotin promoter. The method is useful for the
 CC production of recombinant plants having genes under control of an
 CC osmotin promoter, especially of pathogen inhibiting proteins. Osmotins
 CC are cationic plant proteins, similar to tobacco PR-5-type proteins. The
 CC osmotin genes are under control of hormonal or environmental signals,
 CC including abscisic acid, ethylene, tobacco mosaic virus invention,
 CC salinity, desiccation and wounding. The present sequence represents the
 CC Nicotiana tabacum osmotin protein.
 XX
 XX Sequence 246 AA;
 SQ
 Query Match 71.4%; Score 35; DB 19; Length 246;
 Best Local Similarity 50.0%; Pred. No. 4e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QKYCCSRK 8
 Db 174 QQYCCTQR 181
 I:||||:
 RESULT 11
 ID AAW94510
 AC AAW94510 standard; Protein; 246 AA.
 XX
 AC AAW94510;
 XX
 DT 26-APR-1999 (first entry)
 XX
 DE Tobacco osmotin protein.
 XX
 XX Nicotiana tabacum; tobacco; osmotin; promoter.
 XX
 XX Nicotiana tabacum.
 OS
 XX

PN US5874626-A.
 XX
 PD 23-FEB-1999.
 XX
 PF 12-JAN-1994; 94US-0180428.
 XX
 PR 12-JAN-1994; 94US-0180428.
 PR 20-MAY-1993; 93US-0065147.
 XX
 PA (PURD) PURDUE RES FOUND.
 XX
 XX Bressan R, Hasegawa PM;
 PI WPI; 1999-180080/15.
 DR N-PSDB; AAX16340.
 XX
 XX Tobacco osmotin gene promoter - and recombinant construct comprising
 PT foreign gene under its control
 PT
 PS Disclosure; Fig 9A; 26pp; English.
 XX
 XX The present sequence represents the Nicotiana tabacum (tobacco) osmotin
 CC protein. The present invention specifically claims a DNA fragment
 CC comprising the osmotin gene promoter sequence with no more than 100 bp
 CC of the coding sequence. Also described are methods which are useful in
 CC processes for achieving expression of a coding sequence from a foreign
 CC gene in a host cell.
 XX
 XX Sequence 246 AA;
 SQ
 Query Match 71.4%; Score 35; DB 20; Length 246;
 Best Local Similarity 50.0%; Pred. No. 4e+02;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QKYCCSRK 8
 Db 174 QQYCCTQR 181
 I:||||:
 RESULT 12
 ID ABG09140
 AC ABG09140 standard; Protein; 316 AA.
 XX
 AC ABG09140;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #9131.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR N-PSDB; AAS73327.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess

PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 39499; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 316 AA;
 Query Match 71.4%; Score 35; DB 22; Length 316;
 Best Local Similarity 83.3%; Pred. No. 4.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 3 YCCSRK 8
 Db 276 FCCSRK 281
 RESULT 13
 AAB25483
 ID AAB25483 standard; Protein; 345 AA.
 XX
 AC AAB25483;
 XX
 DT 27-NOV-2000 (first entry)
 XX
 DE Eucalyptus grandis cell signalling involved protein SEQ ID NO:802.
 XX
 KW Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
 KW plant cell signalling; modulation; transgenic plant; pathogen; growth;
 KW environmental change; development; cell proliferation; differentiation;
 KW elongation; survival; disease resistance; nutrient metabolism.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200042171-A1.
 XX
 PD 20-JUL-2000.
 XX
 PF 11-JAN-2000; 2000WO-US00724.
 XX
 PR 12-JAN-1999; 99US-0228986.
 PR 01-NOV-1999; 99US-0162866.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Strabala TJ, Nieuwenhuizen NJ;
 XX
 DR WPI; 2000-476052/41.
 XX
 XX Isolated polynucleotide encoding a polypeptide involved in cell
 PT signaling used for generating transgenic plants with modified responses
 PT to external signals -
 XX

PS Claim 3; Page 371-372; 527pp; English.
 XX
 CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
 CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
 CC pine (Pinus radiata also known as Monterey pine). The protein sequences
 CC are involved in cell signalling. The polynucleotide and protein
 CC sequences can be used to modify the response of plant cells to external
 CC signals e.g. environmental changes or pathogens during the growth and
 CC development of a plant. They can be used to modify cell proliferation,
 CC differentiation, elongation and survival, resistance to disease and
 CC nutrient metabolism. Examples of modifications which can be produced are
 CC altered fruit ripening and senescence of leaves and flowers e.g. to
 CC delay senescence and prolong the life of cut flowers or enhance
 CC senescence of reproductive organs to engineer sterile plants. Other
 CC modifications can be used to delay senescence in selected cell types or
 CC organs providing fruit and vegetables which have a longer shelf life
 CC between harvest and consumption, or to decrease branching frequency in
 CC forest tree species giving long stretches of valuable knot-free clear
 CC wood which can be used in solid timber furniture and veneers.
 XX
 SQ Sequence 345 AA;
 Query Match 71.4%; Score 35; DB 21; Length 345;
 Best Local Similarity 83.3%; Pred. No. 5.2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 3 YCCSRK 8
 Db 306 YCCSRK 311
 RESULT 14
 AAR84882
 ID AAR84882 standard; Protein; 2414 AA.
 XX
 AC AAR84882;
 XX
 DT 01-FEB-1996 (first entry)
 XX
 DE Transcription factor p300.
 XX
 KW Transcription factor; p300; adenovirus; early region 1A; E1A;
 KW cancer; diagnosis.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 11..17
 FT /label= Nuclear_location_signal
 FT Region 342..421
 FT /label= C/H-rich_region_1
 FT /note= "cysteine/histidine-rich region containing
 FT 2 putative zinc finger motifs"
 FT Domain 1070..1134
 FT /label= Bromodomain
 FT Region 1162..1461
 FT /label= C/H-rich_region_2
 FT Region 1622..1821
 FT /label= C/H-rich_region_3
 XX
 XX WO9528499-A1.
 XX
 XX 26-OCT-1995.
 XX
 XX 13-APR-1995; 95WO-US04682.
 XX
 XX 14-APR-1994; 94US-0227536.
 XX
 XX (DAND) DANA FARBER CANCER INST INC.
 XX Eckner R, Ewen M, Livingston D;
 XX
 XX WPI; 1995-373813/48.
 DR

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:08 ; Search time 53.7749 Seconds
(without alignments)
952.238 Million cell updates/sec

Title: US-09-606-129a-1

Perfect score: 1508

Sequence: 1 MNAEPERKFGVVVGVGRAG.....RIHLCLGLAEIQKYCCSRK 296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1494	99.1	296	4 Q96QL4	Q96ql4 homo sapien
2	1484	98.4	296	4 Q9BRW8	Q9brw8 homo sapien
3	1253.5	83.1	295	11 Q9CY64	Q9cy64 mus musculus
4	1253.5	83.1	303	11 Q9DD21	Q9dd21 mus musculus
5	450	29.8	92	4 Q95019	Q95019 homo sapien
6	218.5	14.5	328	16 P72782	P72782 synecocyst
7	155.5	10.3	371	17 Q970H0	Q970h0 sulfolobus
8	142.5	9.4	382	16 Q986T3	Q986t3 rhizobium 1
9	142	9.4	318	17 Q26961	Q26961 methanother
10	139	9.2	393	16 P94437	P94437 bacillus su
11	133	8.8	360	16 Q9W254	Q9w254 thermotoga
12	132.5	8.8	364	10 Q23580	Q23580 arabidopsis
13	131.5	8.7	324	16 Q9CF67	Q9cf67 lactococcus
14	130.5	8.7	351	16 Q98DB4	Q98db4 rhizobium 1
15	129	8.6	335	5 Q9VQB3	Q9vqb3 drosophila
16	128.5	8.5	368	3 Q42896	Q42896 schizosacch

17	128.5	8.5	369	3 Q13991	Q13991 schizosacch
18	128	8.5	319	2 Q93PS4	Q93ps4 comamonas t
19	124.5	8.3	315	2 Q9KWL3	Q9kwl3 pseudomonas
20	122.5	8.1	323	16 Q92F05	Q92f05 listeria in
21	122	8.1	338	16 Q92ET3	Q92et3 listeria in
22	121.5	8.1	346	16 Q99X04	Q99x04 staphylococ
23	121.5	8.1	371	17 Q95939	Q95939 pyrococcus
24	120.5	8.0	369	2 Q9L307	Q9l307 streptomyce
25	119	7.9	341	17 Q9UXX4	Q9uxx4 pyrococcus
26	117	7.8	328	2 Q9LCW9	Q9lcw9 streptomyce
27	116.5	7.7	342	2 Q9ZA33	Q9za33 streptomyce
28	116	7.7	330	2 Q52566	Q52566 amycolatops
29	116	7.7	343	16 Q926S4	Q926s4 listeria in
30	116	7.7	346	2 Q54531	Q54531 streptomyce
31	116	7.7	358	2 Q52495	Q52495 listeria mo
32	115	7.6	358	16 Q929K9	Q929k9 listeria in
33	114.5	7.6	308	16 Q920Y5	Q920y5 rhizobium m
34	114	7.6	360	16 Q9HU35	Q9hu35 pseudomonas
35	113.5	7.5	305	16 Q9CELA	Q9cel4 lactococcus
36	113.5	7.5	349	16 Q929L3	Q929l3 listeria in
37	113	7.5	334	16 Q9WYP5	Q9wyp5 thermotoga
38	113	7.5	378	2 Q9RDH5	Q9rdh5 streptomyce
39	112.5	7.5	357	16 Q9CIU1	Q9ciul lactococcus
40	111.5	7.4	331	16 Q98GV0	Q98gv0 rhizobium 1
41	110	7.3	314	2 Q93N64	Q93n64 coxiella bu
42	109	7.2	321	2 Q9ZGC3	Q9zgc3 streptomyce
43	109	7.2	367	2 Q9F924	Q9f924 pasteurella
44	109	7.2	367	16 Q9CK74	Q9ck74 pasteurella
45	109	7.2	371	2 Q9RK99	Q9rk99 streptomyce

ALIGNMENTS

RESULT 1

Q96QL4 ID Q96QL4 PRELIMINARY; PRT; 296 AA.
AC Q96QL4;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:14706).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008456; AAH08456.1; -.
SQ SEQUENCE 296 AA; 33428 MW; 2CF2AA7F1CDB9707 CRC64;

Query Match	99.1%	Score 1494;	DB 4;	Length 296;
Best Local Similarity	99.0%	Pred. No. 5.8e-116;		
Matches 293;	Conservative 2;	Mismatches 1;	Indels 0;	Gaps 0;
Qy 1	MNAEPERKFGVVVGVGRAGSVYMRDLRNPFPSSAFNLIGFVSRRELGSIDGVQQISLE 60			
Db 1	MNAEPERKFGVVVGVGRAGSVYMRDLRNPFPSSAFNLIGFVSRRELGSIDGVQQISLE 60			
Qy 61	DALSSQEVAVYICSESSSHEDYIRQFLNAGKHVLEVPMTLSLAAQELWELAEQKGV 120			
Db 61	DALSSQEVAVYICSESSSHEDYIRQFLNAGKHVLEVPMTLSLAAQELWELAEQKGV 120			
Qy 121	LHEERVELLMEEFAFKKEVVGKDLLKGLFTSDPLEEDREGFPFAGSGISRLTWLVSLF 180			
Db 121	LHEERVELLMEEFAFKKEVVGKDLLKGLFTSDPLEEDREGFPFAGSGISRLTWLVSLF 180			
Qy 181	GELSLVSATLEERKEDQYMKMTVCLETETKKSPLSWIEEKGPGLKRNRYLSFHFKSGSLN 240			
Db 181	GELSLVSATLEERKEDQYMKMTVCLETETKKSPLSWIEEKGPGLKRNRYLSFHFKSGSLN 240			

```
Qy 241 VPNVGNKNIFLKQDNIFVQKLLGFSEKELAAEKKRILHCLGLAEETQKYCCSRK 296
|||||
Db 241 VPNVGNKNIFLKQDNIFVQKLLGFSEKELAAEKKRILHCLGLAEETQKYCCSRK 296

RESULT 2
Q9BRW8
ID AC Q9BRW8 PRELIMINARY; PRT; 296 AA.
AC Q9BRW8:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO BLIVERDIN REDUCTASE A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=BRAIN, PRIMITIVE NEUROECTODERMAL;
RA Strausberg R.;
RL Submitted (apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005902; AA05902.1; -.
DR InterPro; IPR000683; GFO_IDH_MocA.
DR Pfam; PF01408; GFO_IDH_MocA; 1.
SQ SEQUENCE 296 AA; 33432 MW; A389AFDDB6ADB6F CRC64;

Query Match 98.48; Score 1484; DB 4; Length 296;
Best Local Similarity 98.3%; Pred. No. 3.9e-115;
Matches 291; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MNAEPERKFGVVVGVGRAGSVRMRLNPHPSAFLNLGIFVSRRELGLSDGVQOISLE 60
|||
Db 1 MNTPEKFGVVVGVGRAGSVRMRLNPHPSAFLNLGIFVSRRELGLSDGVQOISLE 60

Qy 61 DALSSQEVAVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120
|||||
Db 61 DALSSQEVAVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120

Qy 121 LHEEHVELLMEEFAFLKKEVGVKDLLGSLFTSDPLEEDRFGPAPSGISRLTWLSLF 180
|||||
Db 121 SHEEHVELLMEEFAFLKKEVGVKDLLGSLFTAGPLEEERFGPAPSGISRLTWLSLF 180

Qy 181 GELSLSATLEERKEDQYMKTVCTLETKKSPLSWIEKPGGLKRNRYLSFHFKSGSLEN 240
|||||
Db 181 GELSLSATLEERKEDQYMKTVCTLETKKSPLSWIEKPGGLKRNRYLSFHFKSGSLEN 240

Qy 241 VPNVGNKNIFLKQDNIFVQKLLGFSEKELAAEKKRILHCLGLAEETQKYCCSRK 296
|||||
Db 241 VPNVGNKNIFLKQDNIFVQKLLGFSEKELAAEKKRILHCLGLAEETQKYCCSRK 296
```

```
RESULT 3
Q9CY64
ID AC Q9CY64 PRELIMINARY; PRT; 295 AA.
AC Q9CY64:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 2500001N03RIK PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
```

```
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momabets P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010847; BAB27219.1; -.
DR MGD; MGI:1917355; 2500001N03RIK.
DR InterPro; IPR000683; GFO_IDH_MocA.
DR Pfam; PF01408; GFO_IDH_MocA; 1.
SQ SEQUENCE 295 AA; 33524 MW; F2E1682BD77032A4 CRC64;

Query Match 83.1%; Score 1253.5; DB 11; Length 295;
Best Local Similarity 82.9%; Pred. No. 5.1e-96;
Matches 242; Conservative 28; Mismatches 21; Indels 1; Gaps 1;

Qy 1 MNAEPERKFGVVVGVGRAGSVRMRLNPHPSAFLNLGIFVSRRELGLSDGVQOISLE 60
|||
Db 1 MSTEPEKFGVVVGVGRAGSVRIRDLKDPH-SFAFLNLGIVSRRELGLSDNVOISLE 59

Qy 61 DALSSQEVAVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120
|||||
Db 60 DALRSQEVAVAYICTESSSHEDYIRQFLQAGKHVLYEYPMALSPFAAAQELWELAAQGRV 119

Qy 121 LHEEHVELLMEEFAFLKKEVGVKDLLGSLFTSDPLEEDRFGPAPSGISRLTWLSLF 180
|||||
Db 120 LHEEHVELLMEEFAFLKKEVGVKDLLGSLRFTASPLEEEKFGPAPSGISRLTWLSLF 179

Qy 181 GELSLSATLEERKEDQYMKTVCTLETKKSPLSWIEKPGGLKRNRYLSFHFKSGSLEN 240
|||||
Db 180 GELSLSATLEERKEDQYMKTVCTLETKKSPLSWIEKPGGLKRNRYLSFHFKSGSLEN 239

Qy 241 VPNVGNKNIFLKQDNIFVQKLLGFSEKELAAEKKRILHCLGLAEETQKYC 292
|||||
Db 240 VPNVGNKNIFLKQDDIFQKLLGQVSAEDLAEKKRILHCLGLASDIQRLC 291

RESULT 4
Q9DD21
ID AC Q9DD21 PRELIMINARY; PRT; 303 AA.
AC Q9DD21:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 0610006A11RIK PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
```

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., BarSh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyok-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK002231; BAB21950.1; -;
 DR MGD; MGI:1915580; 0610006A11Rik.
 DR InterPro; IPR000683; GFO_IDH_MoCA.
 DR Pfam; PF01408; GFO_IDH_MoCA; 1.
 SQ SEQUENCE 303 AA; 34491 MW; 52D8A3B02EE956EB CRC64;

Query Match 83.1%; Score 1253.5; DB 11; Length 303;
 Best Local Similarity 82.9%; Pred. No. 5.3e-96;
 Matches 242; Conservative 26; Mismatches 21; Indels 1; Gaps 1;
 QY 1 MNAEPERKFGVVVGVGRAGSVNRDLRNPSPSAFLNLGFSVRRELGSIDGVOQISLE 60
 DB 1 MSTEPRKRKFGVVVGVGRAGSVNRDLKDPH-SSAFLNLGFSVRRELGSIDNVRQISLE 59
 QY 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLVEYPMPTLSLAAQELWELAEOKGV 120
 DB 60 DALRSQVEDVAICETSSSHEDYIRQFLQAGKHVLVEYPMALSFPAQELWELAAQGRV 119
 QY 121 LHEHVELLMEEFAFLKKEVVGKDLLKGSLLFTSDPLEEDRFGEFPAPFSGISRLTWLSLF 180
 DB 120 LHEHVELLMEEFAFLKKEVVGKDLLKGSLLFTSDPLEEDRFGEFPAPFSGISRLTWLSLF 179
 QY 181 GELSLSVATLEERKEDQYMKTVCTEKKSPLSWIEKPGKLRNRYLSHFHKSGLSEN 240
 DB 180 GELSLSATMENRERDQYMKTVCTEKKSPLSWIEKPGKLRNRYLSHFHKSGLSEN 239
 QY 241 VPNGVKNKIFLKDNITFVKLLQGFSEKELAAEKRLHLGLAEETOKYC 292
 DB 240 VPNGVKNKIFLKDNITFVKLLQGFSEKELAAEKRLHLGLAEETOKYC 291

RESULT 5
 O95019 PRELIMINARY; PRT; 92 AA.
 AC O95019;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE BILIVERDIN-IX ALPHA REDUCTASE (FRAGMENT).
 GN BLVRA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cordes M., Wohlman P., Morris M.,
 RT "The sequence of Homo sapiens PAC clone RP5-974M14.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Waterston R.,
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004939; AAD05025.1;
 FT NON_TER 1
 FT NON_TER 92
 SQ SEQUENCE 92 AA; 10552 MW; 39033110849BD93 CRC64;
 Query Match 29.8%; Score 450; DB 4; Length 92;
 Best Local Similarity 96.7%; Pred. No. 4.5e-30;
 Matches 89; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 119 KVLHEHVELLMEEFAFLKKEVVGKDLLKGSLLFTSDPLEEDRFGEFPAPFSGISRLTWLS 178
 DB 1 KVLHEHVELLMEEFAFLKKEVVGKDLLKGSLLFTAGPLEEERFGFPAPFSGISRLTWLS 60
 QY 179 LFGELSLSVATLEERKEDQYMKTVCTEKK 210
 DB 61 LFGELSLSVATLEERKEDQYMKTVCTEKK 92
 RESULT 6
 P72782 PRELIMINARY; PRT; 328 AA.
 AC P72782;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE BILIVERDIN REDUCTASE.
 GN SLR1784.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuoka A., Muraki A., Nakazaki N., Nardou K., Okumura S.,
 RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.,
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D90900; BAA16797.1; -;
 DR InterPro; IPR000683; GFO_IDH_MoCA.
 DR InterPro; IPR004104; GFO_IDH_MoCA_C.
 DR Pfam; PF01408; GFO_IDH_MoCA; 1.
 DR Pfam; PF02894; GFO_IDH_MoCA_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 328 AA; 36640 MW; 51474ABD26B2C7EF CRC64;

Query Match 14.5%; Score 218.5; DB 16; Length 328;
 Best Local Similarity 25.2%; Pred. No. 4.2e-10;
 Matches 82; Conservative 57; Mismatches 141; Indels 45; Gaps 12;
 QY 3 AEPRKFGVVVGVGRAGSVNRDLRNPSPSAFLNLGFSVRRELGSID-----GVO-Q 56
 DB 8 ATPVR---VGIVGTGYAAQRAAEVFRGDRS---QLVSFWGNSANTAKFADTFGVRPQ 60
 QY 57 ISLEDALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLVEYPMPTLSLAAQELWELAEQ 116
 DB 61 OSWALINDPEIDLVIATINQLHGAIAEALQAGKHVVLEYPALTYAMGKKLOQLARE 120
 QY 117 KGKVLHEHVELLMEEFAFLKKEV---VGKDL-LKGSLLFTSDPLEE-----DRFGFPAP 167
 DB 121 KGKLLHVEHIELLGGVHQAIRQLCKIGEVFARYSTIMGQNPAPQRTYHHQGFGLPV 180
 QY 168 SGISRLTWLSLFGELSLSVATLE---ERKEDQYMK---MTVCLETEKKSPLSWIEKPG 222
 DB 181 AALSRSRFTDLFGTVQVQVDAQCRFDQNPPEYFRACLATAYLQFNNGLKAEIVYKGEV 240
 QY 223 LKRN-RYLSFHFKSGSL-----ENVPNVGVKNKIFLKDNITFVKLLQGF 267
 DB 241 PHONERIFTLHGRDGTGLIFVGTGRLIOGQTETETITVGSRRGLFRQDTEAVLDYLT---T 297
 QY 268 EKELAAEKRLHLGLAEETOKYC 292
 DB 298 GKPLYVDLEASLYALEVADLCAQAC 322
 RESULT 7
 Q97UHO

```
ID Q97UHO PRELIMINARY; PRT; 371 AA.
AC Q97UHO;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DEHYDROGENASE, PUTATIVE.
GN SSO3049.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11477726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
  Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
  De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
  RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
  Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
  RA Charlebois R.L., Doolittle W.F., Duquet M., Gaasterland T.,
  RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
  RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
  RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006896; AA43149.1; -.
DR InterPro; IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
DR Pfam; PF02894; GFO_IDH_MOCA_C; 1.
KW Complete proteome.
SQ SEQUENCE 371 AA; 41760 MW; DAAE16466D53306B CRC64;

Query Match 10.3%; Score 155.5; DB 17; Length 371;
Best Local Similarity 20.6%; Pred. No. 8.3e-05;
Matches 66; Conservative 60; Mismatches 115; Indels 79; Gaps 12;

QY 7 RKFGVVVGVGRAGSVRMRLRNPHPSAFNLNIGFVSRLGSDGV-QQISLE 60
   : : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 KRIGAVVVLGSGIKTHVKALKDLEKETEFVKLVAVVDQIKAEIKSGEYGTPTTID 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 DALSSQEVVAYICSSSHEDYIRQFLNAGKHVLPWMTLSLAAAGLWELAEQK--- 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 EVLRNSEVDVISTATPSYLHAPQAILAEYKGVIVKPMATTLIAGAKEMVSRAEKRV 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 -GKVLHEHVELLMEEFAFLKKEVVGKDL-----LKGSLLTSDPL----- 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 LGVIFQERYA----PDIRRLKNDIL-RELGRYLIASELKWYRDMKG--YKRD EIA 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 ----BEDRFGPAPSGSRITWLVSLFGLSELVSATLEERKEDQYMKWVLCLETKK 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 RGMWNTGGGVMTNQGIHTIDLMIWLNAGEEVSFGVDNLTHDG-----IE----- 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 214 SWTEKGPGLKRNRYLSFHFKSGSLENNPVN-----GVKNKIFLKDONI 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 --VEDTAVAIWR-----YKNGALGTISQTVSMKPTTYQYKIRVNGSNGFVEITD 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 FVQKLLGFSEKELAEKKR 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 STVAIEGKIEKSSVEYKK 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
Q986T3 PRELIMINARY; PRT; 382 AA.
ID Q986T3;
AC Q986T3;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE DEHYDROGENASE.
GN MLL7219.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
  RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
  RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
  RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
  RA Takeuchi C., Yamada M., Tabata S.;
  RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
  RT Mesorhizobium loti.";
  RL DNA Res. 7:331-338(2000).
  DR EMBL; AP003011; BAB53370.1; -.
  DR InterPro; IPR000683; GFO_IDH_MOCA.
  DR Pfam; PF01408; GFO_IDH_MOCA; 1.
  KW Complete proteome.
  SQ SEQUENCE 382 AA; 40593 MW; 785DA27E3D3308C5 CRC64;

Query Match 9.4%; Score 142.5; DB 16; Length 382;
Best Local Similarity 24.1%; Pred. No. 0.001;
Matches 64; Conservative 39; Mismatches 104; Indels 59; Gaps 9;

QY 1 MNAPERKFGVVVGVGRAGSVRMRLRNPHPSAFNLNIGFVSRLGSD----- 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MQAKSEAKLAIGVIG---CGNISMTYLRN-----AALFGGIELRACADISADMAALRA 50
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 53 --GVQOISLEDALSSQEVVAYICSSSHEDYIRQFLNAGKHVLPWMTLSLAAAGL 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51 KEYGIQALGVDAALLSDPEIDLVLNLTIPAAHFDLSLSAGKHKVFTKEKPLATSGDGR 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 LWELAEQKGVLRHE-----EHVELLMEEFAFLKKEVVGKDLKGLSLTSDPLEE 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 LVAEAAKRGVLLGSAPDTFLGAAGRRARRRLMDEGA-IGRPVGTGTFPMGMGRMEHHPNQ 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 DRP---GFPAFS-GISRLTWLVSLFGLSELVSATLEERKEDQYMKWVLCLETKK 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 FYQPGGGVFDMPGYLYLTMLVNLGSPVARVMAMATRGQERL----- 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 IEKGPGLKRNRYLSFHFKSGSLENV 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 ITAEGP-----FKNTTFKVGTPTNV 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
Q26961 PRELIMINARY; PRT; 318 AA.
AC Q26961;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 3-CHLOROBENZATE-3,4-DIOXYGENASE DYHYDROGENASE RELATED PROTEIN.
GN MTH875.
OS Methanothermobacter thermoautotrophicus.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
  RA Alldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
  RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
  RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
  RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
  RA Mcbougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
  RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
  RT "Complete genome sequence of Methanobacterium thermoautotrophicum
  RT deltaH: functional analysis and comparative genomics.";
  RL J. Bacteriol. 179:7135-7155(1997).
  DR EMBL; AE000863; AAB85373.1; -.
  DR InterPro; IPR000683; GFO_IDH_MOCA.
  DR InterPro; IPR004104; GFO_IDH_MOCA_C.
```

DR Pfam; PF01408; GFO_IDH_MOCA; 1.
 DR Pfam; PF02894; GFO_IDH_MOCA_C; 1.
 KW Dioxigenase; Complete proteome.
 SQ SEQUENCE 318 AA; 35653 MW; 3145C2169A98D1E2 CRC64;

Query Match 9.48; Score 142; DB 17; Length 318;
 Best Local Similarity 21.08; Pred. No. 0.00088;
 Matches 57; Conservative 50; Mismatches 96; Indels 68; Gaps 19;

Qy 7 RKFGVVGVGRAGSVRMR---DLRNPSPSSAFNLNIGFVSRRRELGSIDGVQQIISLEDA 63
 Db 2 RQINVGIVGAMGYNHARVYRLKNAN-LMAVSDIMKGTLOKRVANKYDVTGVYVDYENLL 60
 Qy 64 SSOEVEVAYICSSSSSHEDYIROFLNAGKHVLYEVPMTLSLAAQELWELAEORQKVLHE 123
 Db 61 EPIEIEVSVCPVTHVNVMDALEHDKHVLVEKPIAFTLEAEADMVKTARKKGLKGT 120
 Qy 124 EHVE-----LLMEERAFLEKKEVGVKDLKGS-----L 150
 Db 121 GHVERFNPVAVQAKELIENDVIG-DVVSASAKRVGPFPRIKDVGVTIDLAIHDLVNMHY 179
 Qy 151 LFTSDPLEE-----DRFGFPAPFSGISRLTWL-----VSLRGE 182
 Db 180 LF-SEPVAEVAVMGSIIEKECEYEDHAEIMTKFK-SGITGILEVNWLTLPYRRKLAIKGT 237
 Qy 183 LSLVSATLEERKEKDOYKMTVCLETEKKSPL 213
 Db 238 DGIINVDYIDORLDVYGFKAODVDIKHEEPL 268

RESULT 10
 P94437
 ID P94437 PRELIMINARY; PRT; 393 AA.
 AC P94437;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE YFII.
 GN YFII.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AC327.
 RX MEDLINE=97128783; PubMed=8973323;
 RA Yamamoto H., Uchiyama S., Sekiguchi J.;
 RT "The bacillus subtilis chromosome region near 78 degrees contains the
 RT genes encoding a new two-component system, three ABC transporters and
 RT a lipase."
 RL Gene 181:147-151(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Meszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche E., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Tsuchiuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis."
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D78508; BAA11399.1; .
 DR EMBL; Z99108; CAB12657.1; .
 DR InterPro; IPR000683; GFO_IDH_MOCA.
 DR InterPro; IPR004104; GFO_IDH_MOCA_C.
 DR Pfam; PF01408; GFO_IDH_MOCA; 1.
 DR Pfam; PF02894; GFO_IDH_MOCA_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 393 AA; 43584 MW; 3BEE326E1A047602 CRC64;

Query Match 9.28; Score 139; DB 16; Length 393;
 Best Local Similarity 23.38; Pred. No. 0.0021;
 Matches 65; Conservative 51; Mismatches 91; Indels 72; Gaps 12;

Qy 13 VVGVRAGSV---RMRDLRNPSPSAP-LNLIGFVSRRRELGSIDGV----- 54
 Db 17 MVGGRLSQVGYKHRIKALRD---NTAFQLTAGAFDIDAERGKDFGNLGVDAERCYPNY 73
 Qy 55 QOISLEDALSSQVEVAYICSSSSSHEDYIROFLNAGKHVLYEVPMTLSLAAQELWELA 114
 Db 74 QTMFAEAKRQDGEVSVIATPNGTHYBICKAALEAGVHVVICPKPLFTTSAEGOEIKALA 133
 Qy 115 EOKGKVLHEH---VELLMEEFAFLKKEVGVKDLKGLSLFTSDPLEEDRFG----- 164
 Db 134 EKKKIVGTYGFGSGNQLLQNRAMIEQGMIG-DIRVVLDQYT-----HGFCAWDEG 184
 Qy 165 -----PAFSGISRLTWLVSFLGSLVSATLEERKEKDOYKMTVCLETEKKSPL 212
 Db 185 EKISAAQKRWVDPATAGPS-----FVLCDLS-----THYYSQSLIMPMMKIKEL 229
 Qy 213 L-----SWIEERKPGKLRNRYLSFHFKSGSLENVNPVGVN 247
 Db 230 LCDRQSFVGSRAP-LEDNAHLMHYENGAVGTMVTSSIN 267

RESULT 11
 Q9WZ54
 ID Q9WZ54 PRELIMINARY; PRT; 360 AA.
 AC Q9WZ54;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE LIPOPOLYSACCHARIDE BIOSYNTHESIS PROTEIN BPLA.
 GN TM0585.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,

```

RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329 (1999).
DR EMBL; AE001733; AAC35670.1; -.
DR TIGR; TM0585; -.
DR InterPro; IPR000683; GFO_IDH_Moca.
DR InterPro; IPR004104; GFO_IDH_Moca_C.
DR InterPro; IPR003489; Ribosomal_S30.
DR Pfam; PF01408; GFO_IDH_Moca; 1.
DR Pfam; PF02894; GFO_IDH_Moca_C; 1.
DR Complete proteome.
SQ SEQUENCE 360 AA; 41023 MW; 3B17BCD4A9A46EC6 CRC64;

Query Match      8.8%; Score 133; DB 16; Length 360;
Best Local Similarity 23.08; Pred. No. 0.0059;
Matches 65; Conservative 38; Mismatches 103; Indels 76; Gaps 12;

Qy 8 KFGVVVGVGRAGS-----VVRDLRNPHPSAFNLNIGFVSRRELGSID--GVQQ-- 56
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 KLRIALVCGGRIGQKKHVPALIEQDL--PETVAVCDLVEERANRAAEHPEKSGLRPE 58
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 57 --ISLEDAISQOEVEVAYICSESSHEDYIROFLNACKHVLVEYPMTLSLAAQELWELA 114
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 TMTNYRELLKREDVDVISATIESGKHQITMEALEAGKHVLVEKPMALSTKHMNMVVELS 118
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 115 BQKQKVLHEEHVELLMEEFAFLKKEVVGKDLKGLSLFTSPDLEDRFGFPAPFSRLT 174
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 KQKN-----LKLGVFF-----QNRFPNPVQEVRRKKLD 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 175 WLVSDFGELSLVSATLEERKEDQYMKMTVCLETEKKSPLSW--TEEGPGGLKRNRYLSF 232
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 --SAGFGKIFVAVVRNRNEDYKQA-----SHRGTWEMDGGVLMNQ--STH 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 233 -----FKSGSLENPRNVGVNKN---IFLKDQNIFFVKLLG 264
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 AIDLQWFLGGEIEEYGHANTNHPYIEADEGFAIVKFKG 232

RESULT 12
O23580 PRELIMINARY; PRT; 364 AA.
ID O23580
AC O23580;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INOSITOL 2-DEHYDROGENASE HOMOLOG (INOSITOL 2-DEHYDROGENASE LIKE
DE PROTEIN).
GN AT4G17370.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambutt R., Weitzenecker T., Pohl T.M., Terry N.,
RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Leclercq A.,
RA Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Slivey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Millon D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rachman S., Ansoorge W., Cooke R., Berger C.,
RA Delseny M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chalwatzis N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana.";
```

```

RL Nature 391:485-488 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; Z97343; CAB10518.1; -.
DR EMBL; AL161546; CAB78740.1; -.
DR InterPro; IPR000683; GFO_IDH_Moca.
DR InterPro; IPR004104; GFO_IDH_Moca_C.
DR Pfam; PF01408; GFO_IDH_Moca; 1.
DR Pfam; PF02894; GFO_IDH_Moca_C; 1.
DR Complete proteome.
SQ SEQUENCE 364 AA; 40260 MW; B56312BD6CF01B95 CRC64;

Query Match      8.8%; Score 132.5; DB 10; Length 364;
Best Local Similarity 22.7%; Pred. No. 0.0066;
Matches 60; Conservative 50; Mismatches 89; Indels 65; Gaps 13;

Qy 8 KFGVVVGVGRAGSVRMDLR-----NPHSSAFNLN-----IGFVSRRELGSI 51
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 KYGIVGIGMMGREHLNLNHLRLDQGLAVVCTADPHPPSQLAIELAQSGW----ELKVF 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 52 DGVOOISLEDAISQOEVEVAYICSESSHEDYIROFLNACK--HVLVEYPMTLSLAAQOE 109
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 SG-----HEELLSELCDVIWSSPNMTHQILMDIINYSKPHHVLVEKPLCTTTVADCKQ 121
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 110 LWELAEQKGVLEHEHVEL-LMEEFAFLKKEVVGKDLKGLSLFTSPDLEDRFGFPAPFS 168
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 VLEAAKRSRDMVQVGLERYMPPVAKLIEQVGRDP--GNVKMA--IREHREPPFLVKV 177
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 169 GISRL-TW-----LVSLFGLSLVSATLEERKEDQYMKMTVCLETEKKSPLSWI 216
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 GLNKMGLMTLVEKCHCFDLMRLFAGAN-----PVCVMASGMDVNHK 220
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 217 EE-----KPGGLKRNRYLSFHKSG 236
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 DEVYGGKVPDIIDNAYVIIIEFDNG 244

RESULT 13
O9CF67 PRELIMINARY; PRT; 324 AA.
ID O9CF67
AC O9CF67;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE OXIDOREDUCTASE.
GN YPJF.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RS SEQUENCE FROM N.A.
RX STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753 (2001).
DR EMBL; AE006386; AAK05652.1; -.
DR InterPro; IPR000683; GFO_IDH_Moca.
DR Pfam; PF01408; GFO_IDH_Moca; 1.
DR Complete proteome.
SQ SEQUENCE 324 AA; 36603 MW; DBF47D6684EB259C CRC64;

Query Match      8.7%; Score 131.5; DB 16; Length 324;
Best Local Similarity 21.8%; Pred. No. 0.0067;
Matches 65; Conservative 44; Mismatches 88; Indels 101; Gaps 13;

Qy 8 KFGVV-----VVGVRAGSVRMDLRNPHPSAFNLNIGFVSRRELGSDGVQOIS 58
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 KIGIMGAAKIVPRFVAGVKESGQA-----EVTGIARNK-----EKAOKAA 43
```


Search completed: November 1, 2002, 14:49:42
Job time : 56.7749 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:41:32 ; Search time 15.2283 Seconds
(without alignments)
752.611 Million cell updates/sec

Title: US-09-606-129A-1

Perfect score: 1508

Sequence: 1 MNAEPKRGVGVGVGRAG.....RILHCLGLAEIQKYCCSRK 296

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1508	100.0	296	1 BIEA_HUMAN	P53004 homo sapien
2	1245.5	82.6	295	1 BIEA_RAT	P46844 rattus norv
3	152.5	10.1	345	1 YHXX_ECOLI	P46853 escherichia
4	134	8.9	371	1 YH16_SYNY3	P74041 synechocyst
5	118.5	7.9	358	1 YVAA_BACSU	O32223 bacillus su
6	114.5	7.6	328	1 YULF_BACSU	O05265 bacillus su
7	111	7.4	341	1 YRBE_BACSU	P75931 escherichia
8	105.5	7.0	307	1 MVIM_ECOLI	O05184 pseudomonas
9	104	6.9	410	1 PHT4_PSEPU	O44258 comamonas t
10	103.5	6.9	397	1 CBAC_COMTE	P40332 bacillus su
11	103	6.8	342	1 YISS_BACSU	P77376 escherichia
12	103	6.8	346	1 YDGI_ECOLI	O04869 saccharomyc
13	101.5	6.7	349	1 YW94_YEAST	Q07982 zymomonas m
14	101.5	6.7	439	1 GFO_ZYMMO	P47769 mycobacteri
15	100.5	6.7	1316	1 RPOC_MYCTU	O06238 enterococcu
16	98	6.5	988	1 TNP6_ENTFC	P27059 astasia lon
17	97.5	6.5	1076	1 RPOB_ASTLO	P54394 bacillus su
18	97	6.4	931	1 DING_BACSU	O08378 homo sapien
19	96	6.4	579	1 G160_HUMAN	P37168 salmonella
20	95	6.3	307	1 MVIM_SALTY	P49305 rhizobium m
21	95	6.3	334	1 YMO1_RHIME	P02549 homo sapien
22	95	6.3	2418	1 SPCA_HUMAN	Q62209 mus musculu
23	94.5	6.3	993	1 SCPL_MOUSE	P42599 escherichia
24	94	6.2	328	1 YGJR_ECOLI	P12626 azotobacter
25	94	6.2	533	1 ANFA_AZOVI	O67638 aquifex aeo
26	93	6.2	316	1 YH58_AQUAE	Q13439 homo sapien
27	92.5	6.1	2230	1 GOG4_HUMAN	O68965 rhizobium m
28	92	6.1	330	1 M12D_RHIME	P57208 buchnera ap
29	92	6.1	468	1 6PGD_BUCAI	P42499 glycine max
30	92	6.1	1156	1 FLIF_SOYBN	P57175 buchnera ap
31	90.5	6.0	545	1 PHIE_BUCAI	O75330 homo sapien
32	90.5	6.0	724	1 HMMR_HUMAN	P07149 s fatty aci
33	90.5	6.0	2051	1 FAS1_YEAST	

RESULT 1	BIEA_HUMAN	STANDARD;	PRT;	296 AA.
ID	BIEA_HUMAN			
AC	P53004;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Biliverdin reductase A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-reductase).			
DE	BLVRA OR BLVR OR BVR.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=96202961; PubMed=8631357;			
RA	Maines M.D., Polevoda B.V., Huang T.-J., McCoubrey W.K. Jr.;			
RT	"Human biliverdin IXalpha reductase is a zinc-metalloprotein.			
RT	Characterization of purified and Escherichia coli expressed			
RT	enzymes.";			
RL	Eur. J. Biochem. 235:372-381(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Komuro A., Tobe T., Nakano Y., Yamaguchi T., Tomita M.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-117 FROM N.A.			
RA	Cordes M., Wollam C., Carter T.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE OF 3-36; 48-74 AND 228-248.			
RC	TISSUE=Liver;			
RX	MEDLINE=93143333; PubMed=8424666;			
RA	Maines M.D., Trakshel G.M.;			
RT	"Purification and characterization of human biliverdin reductase.";			
RT	Arch. Biochem. Biophys. 300:320-326(1993).			
RN	[5]			
RP	SEQUENCE OF 3-22.			
RC	TISSUE=Liver;			
RX	MEDLINE=95014177; PubMed=7929092;			
RA	Yamaguchi T., Komoda Y., Nakajima H.;			
RT	"Biliverdin-IX alpha reductase and biliverdin-IX beta reductase from			
RT	human liver. Purification and characterization.";			
RL	J. Biol. Chem. 269:24343-24348(1994).			
CC	-/- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN.			
CC	-/- CATALYTIC ACTIVITY: Bilirubin + NAD(P)(+) = Biliverdin + NAD(P)H.			
CC	-/- COFACTOR: BINDS ONE ZINC ION. HAS DUAL PH/COFACTOR (NADH, NADPH)			
CC	-/- SPECIFICITY. USES NADH AT THE ACIDIC PH RANGE (6-6.7) AND NADPH AT			
CC	THE ALKALINE RANGE (8.5-8.7).			
CC	-/- PATHWAY: FINAL STEP IN HEME METABOLISM.			
CC	-/- SUBUNIT: MONOMER.			
CC	-/- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	-/- TISSUE SPECIFICITY: LIVER.			
CC	-/- SIMILARITY: TO E. COLI YHHX.			

34	90.5	6.0	2415	1 SPCA_DROME	P13395 drosophila
35	90	6.0	528	1 UGA3_YEAST	P26370 saccharomyc
36	90	6.0	1679	1 YIO9_YEAST	P40457 saccharomyc
37	90	6.0	3911	1 AKAS2_HUMAN	Q99996 h a-kinase
38	89.5	5.9	386	1 DEGS_BACBR	P54663 bacillus br
39	89.5	5.9	494	1 PRE_STRAG	P13925 streptococc
40	89	5.9	722	1 MFPI1_TOBAC	O9m714 nicotiana t
41	88.5	5.9	792	1 RECG_AQUAE	O67837 aquifex aeo
42	88.5	5.9	1316	1 RPOC_MYCLE	P30761 mycobacteri
43	88.5	5.9	1548	1 SMCY_MOUSE	O62240 mus musculu
44	88	5.8	367	1 YG86_STRPN	Q34748 streptococc
45	87.5	5.8	271	1 PURR_LACLA	O53065 lactococcu

ALIGNMENTS

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X93086; CAA63635.1; -
EMBL; U34877; AAC35588.1; -
DR EMBL; AC005189; AAC25526.1; -
DR MIM; 109750; -
DR InterPro; IPR000683; GFO_IDH_MocA.
DR Pfam; PF01408; GFO_IDH_MocA; 1.
KW Oxidoreductase; NAD; NADP; Zinc.
FT PROPEP 1 2
FT CHAIN 3 296 BILIVERDIN REDUCTASE A.
FT DOMAIN 11 16 POLY-VAL.
FT METAL 280 280 ZINC (POTENTIAL).
FT METAL 281 281 ZINC (POTENTIAL).
FT METAL 292 292 ZINC (POTENTIAL).
FT METAL 293 293 ZINC (POTENTIAL).
FT CONFLICT 3 3 A -> T (IN REF. 2).
FT CONFLICT 154 155 SD -> AG (IN REF. 2).
FT CONFLICT 160 160 D -> E (IN REF. 2).
SQ SEQUENCE 296 AA; 33488 MW; 0DFD3B386F4DFC0A CRC64;

Query Match 100.0%; Score 1508; DB 1; Length 296;
Best Local Similarity 100.0%; Pred. No. 1.2e-108;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNAEPERKFGVVVGVGRAGSVRMRLRNPSPSAFLNLIGFVSRRELGSIDGVQOISLE 60
Db 1 MNAEPERKFGVVVGVGRAGSVRMRLRNPSPSAFLNLIGFVSRRELGSIDGVQOISLE 60

QY 61 DALSSQEVVAYICSSSSSHEDYIROFLNAGKHVLYVEYPMTLSLAAQELWELAEQKGV 120
Db 61 DALSSQEVVAYICSSSSSHEDYIROFLNAGKHVLYVEYPMTLSLAAQELWELAEQKGV 120

QY 121 LHEEHVELLMEEFAFLKKEVGVGKDLKLSLFTSDPLEEDRFPGFAPSGISRLTLVLSLF 180
Db 121 LHEEHVELLMEEFAFLKKEVGVGKDLKLSLFTSDPLEEDRFPGFAPSGISRLTLVLSLF 180

QY 181 GELSLVSATLEERKEQYMKMTVCLETEKKSPLSWIEEKGPKLKNRYLSFHFHKSGLN 240
Db 181 GELSLVSATLEERKEQYMKMTVCLETEKKSPLSWIEEKGPKLKNRYLSFHFHKSGLN 240

QY 241 VPVGVNKNIFLKDQNIQFVKLLGQFSEKELAAEKRRILHCLGLAEIIOKYCCSRK 296
Db 241 VPVGVNKNIFLKDQNIQFVKLLGQFSEKELAAEKRRILHCLGLAEIIOKYCCSRK 296

RESULT 2
ID BIEA_RAT
AC P46844;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Blliverdin reductase A precursor (EC 1.3.1.24) (Blliverdin-IX alpha-reductase).
GN BLVRA OR BLVR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=92156147; PubMed=1371282;
RA Fakhrai H., Mains M.D.;
RT "Expression and characterization of a cDNA for rat kidney blliverdin

reductase. Evidence suggesting the liver and kidney enzymes are the same transcript product."
J. Biol. Chem. 267:4023-4029(1992).
[2]
MUTAGENESIS.
RX MEDLINE=94291657; PubMed=8020496;
RA McCoubrey W.K. Jr., Mains M.D.;
RT "Site-directed mutagenesis of cysteine residues in blliverdin reductase. Roles in substrate and cofactor binding."
Eur. J. Biochem. 222:597-603(1994).
CC -!- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN; DISPLAYS TWO DISTINCT PH OPTIMA USING A DIFFERENT COFACTOR AT EACH PH: NADH AT THE LOWER PH 6.7-6.9 RANGE AND NADPH AT PH 8.5-8.7. NADPH, HOWEVER, IS THE PROBABLE COFACTOR IN BIOLOGICAL SYSTEMS.
CC -!- CATALYTIC ACTIVITY: Bllirubin + NAD(P)(+) = blliverdin + NAD(P)H.
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- PATHWAY: FINAL STEP IN HEME METABOLISM.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: TO E-COLI YHHX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC -----
EMBL; M81681; AAA40830.1; -
DR InterPro; IPR000683; GFO_IDH_MocA.
DR Pfam; PF01408; GFO_IDH_MocA; 1.
KW Oxidoreductase; NAD; NADP; Zinc.
FT PROPEP 1 2
FT CHAIN 3 295 BILIVERDIN REDUCTASE A.
FT DOMAIN 11 16 POLY-VAL.
FT METAL 279 279 ZINC (POTENTIAL).
FT METAL 280 280 ZINC (POTENTIAL).
FT METAL 291 291 ZINC (POTENTIAL).
FT METAL 292 292 ZINC (POTENTIAL).
FT MUTAGEN 73 73 C->A: LOSS OF ACTIVITY.
FT MUTAGEN 280 280 C->A: REDUCED ACTIVITY.
FT MUTAGEN 291 291 C->A: REDUCED ACTIVITY.
SQ SEQUENCE 295 AA; 33565 MW; 219C8EA96C150588 CRC64;

Query Match 82.6%; Score 1245.5; DB 1; Length 295;
Best Local Similarity 82.4%; Pred. No. 1.6e-88;
Matches 244; Conservative 28; Mismatches 23; Indels 1; Gaps 1;

QY 1 MNAEPERKFGVVVGVGRAGSVRMRLRNPSPSAFLNLIGFVSRRELGSIDGVQOISLE 60
Db 1 MDAEPKRKFGVVVGVGRAGSVRLRLDKDPR-SAAFLNLIGFVSRRELGSIDGVQOISLE 59

QY 61 DALSSQEVVAYICSSSSSHEDYIROFLNAGKHVLYVEYPMTLSLAAQELWELAEQKGV 120
Db 60 DALRSQEIDVAYICSSSSSHEDYIROFLQAGKHVLYVEYPMTLSFAAAQELWELAEQKGV 119

QY 121 LHEEHVELLMEEFAFLKKEVGVGKDLKLSLFTSDPLEEDRFPGFAPSGISRLTLVLSLF 180
Db 120 LHEEHVELLMEEFAFLKKEVGVGKDLKLSLFTSDPLEEDRFPGFAPSGISRLTLVLSLF 179

QY 181 GELSLVSATLEERKEQYMKMTVCLETEKKSPLSWIEEKGPKLKNRYLSFHFHKSGLN 240
Db 180 GELSLVSATLEERKEQYMKMTVCLETEKKSPLSWIEEKGPKLKNRYLVNFQFTSGSLEE 239

QY 241 VPVGVNKNIFLKDQNIQFVKLLGQFSEKELAAEKRRILHCLGLAEIIOKYCCSRK 296
Db 240 VPSGVNKNIFLKDQDIFVQKLLDQSAEDLAAEKRRIMHCLGLASDIQKLCQK 295

RESULT 3
YHHX_ECOLI
ID YHHX_ECOLI STANDARD; PRT; 345 AA.

Query Match 6.9%; Score 103.5; DB 1; Length 397;
Best Local Similarity 29.0%; Pred. NO. 0.84;

Query Match	6.8%	Score 103;	DB 1;	Length 342;
Best Local Similarity	25.2%;	pred. No. 0.76;		

Db	226	-----KNKEVIGSFTYGSAGF-----ATEKSVELK	250
<p>RESULT 14</p> <p>GEO_ZYMMO</p> <p>ID GEO_ZYMMO STANDARD; PRT: 439 AA.</p> <p>AC Q07982;</p> <p>DT 01-NOV-1997 (Rel. 35, Created)</p> <p>DT 01-NOV-1997 (Rel. 35, Last sequence update)</p> <p>DT 16-OCT-2001 (Rel. 40, Last annotation update)</p> <p>DE Glucose--fructose oxidoreductase precursor (EC 1.1.99.28) (GFOR).</p> <p>GN Zymomonas mobilis.</p> <p>OS Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;</p> <p>OC Zymomonas.</p> <p>OC NCBI_TaxID=542;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A., AND SEQUENCE OF 53-87.</p> <p>RC STRAIN=ATCC 29191 / ZM6;</p> <p>RX MEDLINE=92165717; PubMed=1537789;</p> <p>RA Kanagasundaram V., Scopes R.K.;</p> <p>RT "Cloning, sequence analysis, and expression of the structural gene encoding glucose-fructose oxidoreductase from Zymomonas mobilis";</p> <p>RL J. Bacteriol. 174:1439-1447(1992).</p> <p>RN [2]</p> <p>RN SEQUENCE OF 1-71 FROM N.A., AND SEQUENCE OF 2-16 AND 53-71.</p> <p>RC STRAIN=ATCC 29191 / ZM6, AND ATCC 31821 / ZM4 / CP4;</p> <p>RX MEDLINE=93321476; PubMed=8472911;</p> <p>RA Loos H., Sahm H., Sprenger G.A.;</p> <p>RT "Glucose-fructose oxidoreductase, a periplasmic enzyme of Zymomonas mobilis, is active in its precursor form.";</p> <p>RL FEMS Microbiol. Lett. 107:293-298(1993).</p> <p>RN [3]</p> <p>RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).</p> <p>RX MEDLINE=97148336; PubMed=8994968;</p> <p>RA Kingston R.L., Scopes R.K., Baker E.N.;</p> <p>RT "The structure of glucose-fructose oxidoreductase from Zymomonas mobilis: an osmoprotective periplasmic enzyme containing non-disociable NADP.";</p> <p>RT Structure 4:1413-1428(1996).</p> <p>RL -!- CATALYTIC ACTIVITY: D-glucose + D-fructose = D-gluconolactone + D-glucitol.</p> <p>CC -!- COFACTOR: CONTAINS ONE TIGHTLY-BOUND NADP(H) MOLECULE PER SUBUNIT.</p> <p>CC -!- PATHWAY: SORBITOL-GLUCONATE PATHWAY.</p> <p>CC -!- SUBUNIT: HOMOTETRAMER.</p> <p>CC -!- SUBCELLULAR LOCATION: Periplasmic.</p> <p>CC -!- SIMILARITY: BELONGS TO THE GEO/IDH/MOCA FAMILY.</p> <p>CC</p> <p>CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).</p> <p>CC -----</p> <p>CC EMBL; M97379; AAA27690.1; .</p> <p>CC EMBL; X73088; CAA51534.1; .</p> <p>CC PDB; 1OFG; 21-APR-97.</p> <p>CC InterPro; IPR000683; GEO_IDH_MocA.</p> <p>CC InterPro; IPR004104; GEO_IDH_MocA_C.</p> <p>CC Pfam; PF01408; GEO_IDH_MocA; 1.</p> <p>CC Pfam; PF02894; GEO_IDH_MocA_C; 1.</p> <p>CC Oxidoreductase; NADP; Signal; Periplasmic; 3D-structure.</p> <p>KW SIGNAL</p> <p>FT CHAIN</p> <p>FT CHAIN 53 439 GLUCOSE--FRUCTOSE OXIDOREDUCTASE.</p> <p>FT CONFLICT 61 61 L -> Q (IN REF. 2; AA SEQUENCE).</p> <p>FT SEQUENCE 439 AA; 47801 MW; C1E4A19F5B34267A CRC64;</p> <p>CC -----</p> <p>CC Query Watch</p> <p>CC Best Local Similarity 25.9%; Score 101.5; DB 1; Length 439;</p> <p>CC Matches 36; Conservative 19; Mismatches 45; Indels 39; Gaps 5;</p>			

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:58 ; Search time 31.4084 Seconds
(without alignments)
905.569 Million cell updates/sec

Title: US-09-606-129a-1

Perfect score: 1508

Sequence: 1 MNAEPERKFGVVGVRAG.....RILHCLGLAEIQRKCCSRK 296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1508	100.0	296	2 S62624	billiverdin reducta
2	1490	98.8	296	2 G02066	billiverdin reducta
3	1245.5	82.6	295	2 A42268	billiverdin reducta
4	221.5	14.7	322	2 A42324	billiverdin reducta
5	218.5	14.5	328	2 S74645	billiverdin reducta
6	155.5	10.3	371	2 F90486	dehydrogenase, pro
7	152.5	10.1	345	2 C65140	hypothetical 38.8
8	150.5	10.0	345	2 AH0994	probable oxidoredu
9	148.5	9.8	345	2 A91165	probable regulator
10	148.5	9.8	345	2 A86011	probable regulator
11	142	9.4	318	1 H69216	3-chlorobenzoate-3
12	139	9.2	393	2 D69803	hypothetical prote
13	136.5	9.1	374	2 AE2337	hypothetical prote
14	134	8.9	371	2 S75555	hypothetical prote
15	133	8.8	338	2 AE1119	oxidoreductase hom
16	133	8.8	360	2 B72359	lipopolysaccharide
17	132.5	8.8	364	2 A71443	probable inositol
18	131.5	8.7	324	2 B86819	oxidoreductase ypj
19	128.5	8.5	368	2 T39296	oxidoreductase - f
20	128.5	8.5	369	2 T38428	hypothetical oxido
21	125.5	8.3	352	2 T42744	hypothetical prote
22	124	8.2	343	2 A11428	oxidoreductases ho
23	123	8.2	356	2 AC2319	hypothetical prote
24	122.5	8.1	323	2 AH1470	oxidoreductase hom
25	122	8.1	338	2 AH1479	oxidoreductase hom
26	121.5	8.1	346	2 F89784	hypothetical prote
27	121.5	8.1	371	1 D71201	hypothetical prote
28	119.5	7.9	323	2 AF1109	oxidoreductase hom
29	119.5	7.9	324	2 AE3650	oxidoreductase (EC

RESULT 1

S62624

billiverdin reductase (EC 1.3.1.24) - human

N;Alternate names: billiverdin IX-alpha reductase

C;Species: Homo sapiens (man)

C;Date: 28-Oct-1996 #sequence.revision 09-May-1997 #text_change 21-Jul-2000

C;Accession: S62624; S62622; S29736

R;Maines, M.D.; Polevoda, B.V.; Huang, T.J.; McCoubrey Jr., W.K.

Eur. J. Biochem. 235, 372-381, 1996

A;Title: Human billiverdin IX-alpha reductase is a zinc-metalloprotein. Characterizati

A;Reference number: S62622; MUID:96202961

A;Accession: S62624

A;Molecule type: mRNA

A;Residues: 1-296 <MAI>

A;Cross-references: EMBL:X93086; NID:gl246748; PIDN:CAA63635.1; PID:gl246749

A;Accession: S62622

A;Molecule type: protein

A;Residues: 3-24,'X',26-27,'X',29-36,48-74;228-234;235-248 <MAF>

R;Maines, M.D.; Frakshel, G.M.

Arch. Biochem. Biophys. 300, 320-326, 1993

A;Title: Purification and characterization of human billiverdin reductase.

A;Reference number: S29736; MUID:93143333

A;Accession: S29736

A;Molecule type: protein

A;Residues: 3-24,'X',26-27,'X',29-36,48-74;228-234;235-248 <MAW>

A;Note: the sequence of peptide 1 from page 323 seems not to belong to this protein

C;Genetics:

A;Gene: BVR

C;Keywords: oxidoreductase

F;3-296/Product: billiverdin reductase IX-alpha #status experimental <MAT>

Query Match 100.0%; Score 1508; DB 2; Length 296;

Best Local Similarity 100.0%; Pred. No. 1.2e-109;

Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNAEPERKFGVVGVRAGSVRMRLNPHPSAFLNLIGFVSRREGSIDGVQOQISLE 60

Db 1 MNAEPERKFGVVGVRAGSVRMRLNPHPSAFLNLIGFVSRREGSIDGVQOQISLE 60

Qy 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLEVPMTLSLAAQELMELAEQKGV 120

Db 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLEVPMTLSLAAQELMELAEQKGV 120

Qy 121 LHEEHVELLMEEFAFLKEVVGVKDLLKGSLLFTSDPLEEDRFGFPFAGISRLTWLSLF 180

Db 121 LHEEHVELLMEEFAFLKEVVGVKDLLKGSLLFTSDPLEEDRFGFPFAGISRLTWLSLF 180

Qy 181 GELSLSVATLEEREDQYMKMTVCLETEKKSPLSWIEKGGCLKRNRYLSFHFSGSLEN 240

Db 181 GELSLSVATLEEREDQYMKMTVCLETEKKSPLSWIEKGGCLKRNRYLSFHFSGSLEN 240

Qy 241 VPNGVNNIFLKDQNIQVQLLGFQSEKELAAEKRLHCLGLAEELQKYCCSRK 296

NADH-dependent dyh
conserved hypothet
hypothetical prote
oxidoreductase (EC
oxidoreductases ho
oxidoreductase hom
oxidoreductase hom
dehydrogenase Atu1
conserved hypothet
myo-inositol 2-deh
myo-inositol 2-deh
D-galactose 1-dehy
conserved hypothet
oxidoreductase ysj
oxidoreductase hom
oxidoreductase hom

Db 241 VPNVGNKNIPLKQDNIFVQKLLGQFSEKELAEKKRILHCLGLAEIQTCCSRK 296
|||||
RESULT 2
G02066
bilverdin reductase (EC 1.3.1.24) - human
N;Alternate names: biliverdin IX-alpha reductase
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C;Accession: G02066
R;Komuro, A.; Tobe, T.; Nakano, Y.; Yamaguchi, T.; Tomita, M.
submitted to the EMBL Data Library, August 1995
A;Reference number: H00768
A;Accession: G02066
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-296 <KOM>
A;Cross-references: EMBL:U34877; NID:g1143231; PID:g1143232
C;Keywords: oxidoreductase

Query Match 98.8%; Score 1490; DB 2; Length 296;
Best Local Similarity 98.6%; Pred. No. 3e-108;
Matches 292; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MNAEPKRFVVVGVGRAGSVRMRLNPHPSAFNLIGFVSRRELGSIDGVOQISLE 60
|||
Db 1 MNTPEPKRFVVVGVGRAGSVRMRLNPHPSAFNLIGFVSRRELGSIDGVOQISLE 60
|||||

Qy 61 DALSSQEVAVAYICSSSSHHEDYIRQFLNAGKHVLYVEYPMPTLSLAAQELWELAEQKGV 120
|||||
Db 61 DALSSQEVAVAYICSSSSHHEDYIRQFLNAGKHVLYVEYPMPTLSLAAQELWELAEQKGV 120
|||||

Qy 121 LHEEHVELLMEEFAFLKKEVVGKDLLGSLFTSDPLEEDRFGFPAPSGISRLTWLSLF 180
|||||
Db 121 LHEEHVELLMEEFAFLKKEVVGKDLLGSLFTSDPLEEDRFGFPAPSGISRLTWLSLF 180
|||||

Qy 181 GELSIVSATLEEREDQYMKTVCTLETKKSPLSWIEKPGKLRNRYLSFHFKSGSLEN 240
|||||
Db 181 GELSIVSATLEEREDQYMKTVCTLETKKSPLSWIEKPGKLRNRYLSFHFKSGSLEN 240
|||||

Qy 241 VPNVGNKNIPLKQDNIFVQKLLGQFSEKELAEKKRILHCLGLAEIQTCCSRK 296
|||||
Db 241 VPNVGNKNIPLKQDNIFVQKLLGQFSEKELAEKKRILHCLGLAEIQTCCSRK 296
|||||

RESULT 3
A42268
bilverdin reductase (EC 1.3.1.24) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A42268
R;Fakhrai, H.; Maines, M.D.
J. Biol. Chem. 267, 4023-4029, 1992
A;Title: Expression and characterization of a cDNA for rat kidney biliverdin reductase.
A;Reference number: A42268; MUID:92156147
A;Accession: A42268
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid; protein
A;Residues: 1-295 <FAK>
A;Cross-references: GB:M81681; NID:g203177; PIDN:AAAA0830.1; PID:g203178
A;Experimental source: kidney
A;Note: sequence extracted from NCBI backbone (NCBIP:82800)
C;Keywords: liver; oxidoreductase

Query Match 82.6%; Score 1245.5; DB 2; Length 295;
Best Local Similarity 82.4%; Pred. No. 2.8e-89;
Matches 244; Conservative 28; Mismatches 23; Indels 1; Gaps 1;

Qy 1 MNAEPKRFVVVGVGRAGSVRMRLNPHPSAFNLIGFVSRRELGSIDGVOQISLE 60
|||||
Db 1 MNAEPKRFVVVGVGRAGSVRLRLDKDPR-SAAFNLIGFVSRRELGSIDVQRISLE 59
|||||

Qy 61 DALSSQEVAVAYICSSSSHHEDYIRQFLNAGKHVLYVEYPMPTLSLAAQELWELAEQKGV 120
|||||
Db 60 DALRSQEDIVAY ICSSSSHHEDYIRQFLQAGKHVLYVEYPMPTLSFAAAQELWELAAQGRV 119
|||||

Qy 121 LHEEHVELLMEEFAFLKKEVVGKDLLGSLFTSDPLEEDRFGFPAPSGISRLTWLSLF 180
|||||
Db 120 LHEEHVELLMEEFEFLRREVLGKELLGSLRFTASPLEEERFGFPAPSGISRLTWLSLF 179
|||||

Qy 181 GELSIVSATLEEREDQYMKTVCTLETKKSPLSWIEKPGKLRNRYLSFHFKSGSLEN 240
|||||
Db 180 GELSIVSATLEEREDQYMKTVCTLETKKSPLSWIEKPGKLRNRYVNFQFTSGSLEE 239
|||||

Qy 241 VPNVGNKNIPLKQDNIFVQKLLGQFSEKELAEKKRILHCLGLAEIQTCCSRK 296
|||||
Db 240 VPSVGNKNIPLKQDNIFVQKLLDQVSAEDLAEKKRIMHCLGLASDIQKLCHQKK 295
|||||

RESULT 4
AF2324
bilverdin reductase [imported] - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
A;Note: Anabaena sp. (Strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C;Accession: AF2324
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriku
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
Anabaena sp. PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AF2324
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-322 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA075848.1; PID:g17133284; GSPDB:GN001179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: bvdR

Query Match 14.7%; Score 221.5; DB 2; Length 322;
Best Local Similarity 26.0%; Pred. No. 1e-09;
Matches 87; Conservative 53; Mismatches 109; Indels 85; Gaps 15;

Qy 13 VVGVRAGSVRMRLNPHPSAFNLIGFVSRRELGSIDGVOQ 56
|||
Db 3 IVGTGYAKLR-----AAFLDESHLVAIAGSKLERTQTFAQAYQAEVITQWQ 53
|||

Qy 57 ISLEDALSSQEVAVAYICSSSSHHEDYIRQFLNAGKHVLYVEYPMPTLSLAAQELWELAEQ 116
|||
Db 54 L-----VEREDIDLVICTINRHDGAIARAALTAGKHVIVEYPLSVDLTEAEELIALAKT 108
|||

Qy 117 KGKVLHEEHVELLMEEFAFLKKEVVGKDLLGSLIF-----TSDP-----LEEDRF 162
|||
Db 109 QOKLLHVEHIELLGLHQAALK-----QNLDKVGHFLVRYSTINPQNAPRKYTYNHAMF 163
|||

Qy 163 GFPAFSGISRLTWLSLFGELSVASLLEERK---EDQYMKMTVCTLETKKSPLSW--- 215
|||
Db 164 GFPLIGALSRLRLTDLFGKVFVNC---HQRYWEIEPEYQTCFCM-----TOLCFTSGL 216
|||

Qy 216 ----IEKPGGL-KRNRYLSFHFKSGSL-----ENVP-NVGWNKNIFLKDQN 256
|||
Db 217 LAQVIYKGESLWQPERKFEVHGDNALIFDGDGTFFIQSGESTPVELGTRRGLFAKDT 276
|||

Qy 257 IFVOKLLGQFSEKELAEKKRILHCLGLAEIQTCCSRK 290
|||
Db 277 MVLDDHI---FDGAPLYVTPPEESLYTLRVADAQR 307
|||

RESULT 5
S74645
bilverdin reductase (EC 1.3.1.24) - Synechocystis sp. (strain PCC 6803)
N;Alternate names: hypothetical protein slr1784
C;Species: Synechocystis sp.
A;Variety: PCC 6803

Db 237 LANDVPCQINITSAHGG---RGHWLEIYGKCSL 268

Db 237 LANDVPCQINITSAHGG---RGHWLEIYGKCSL 268

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run On: November 1, 2002, 14:44:48 ; Search time 24.746 Seconds
(without alignments)
292.168 Million cell updates/sec

Title: us-09-606-129a-1

Perfect score: 1508

Sequence: 1 MNAEPKFKGVVGVGVRAG.....RILHCLGLAEIOKYCCSRK 296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues .

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124.5	8.3	315	4	US-09-720-817-3
2	124.5	8.3	315	4	US-09-720-817-6
3	122.5	8.1	280	4	US-09-720-817-1
4	108	7.2	229	4	US-08-936-165A-519
5	107	7.1	332	4	US-09-036-987A-15
6	107	7.1	332	4	US-09-370-700-15
7	98	6.5	988	2	US-08-286-819A-19
8	98	6.5	988	2	US-08-980-357-19
9	90.5	6.0	351	1	US-08-402-217A-2
10	90.5	6.0	351	1	US-08-700-178-2
11	90.5	6.0	351	3	US-08-995-654-2
12	88	5.8	551	3	US-08-886-886-15
13	87	5.8	264	2	US-07-857-224B-9
14	86	5.7	759	1	US-08-676-967-1
15	86	5.7	759	1	US-08-676-974-1
16	86	5.7	759	2	US-09-098-487-1
17	84	5.6	487	2	US-08-712-709-8
18	84	5.6	487	3	US-09-111-444-8
19	84	5.6	487	4	US-09-541-228-8
20	83.5	5.5	514	2	US-08-960-022-14
21	82	5.4	270	2	US-08-852-743-5
22	82	5.4	270	3	US-09-185-370-5
23	82	5.4	451	2	US-08-372-652-2
24	82	5.4	451	5	PCT-US95-16311-2
25	82	5.4	484	2	US-08-372-652-1
26	82	5.4	484	5	PCT-US95-16311-1
27	82	5.4	1205	4	US-09-330-330-1

28	81.5	5.4	323	1	US-07-992-827D-1	Sequence 1, Appli
29	81.5	5.4	323	1	US-08-216-593-1	Sequence 1, Appli
30	81.5	5.4	323	5	PCT-US93-12380-1	Sequence 1, Appli
31	81.5	5.4	988	3	US-08-851-843A-69	Sequence 69, Appli
32	81.5	5.4	988	4	US-08-974-349A-112	Sequence 112, App
33	81.5	5.4	988	4	US-08-854-050-69	Sequence 69, Appli
34	81.5	5.4	988	4	US-09-430-323-69	Sequence 69, Appli
35	81	5.4	1261	3	US-09-080-855-2	Sequence 2, Appli
36	80.5	5.3	539	3	US-08-557-210A-5	Sequence 5, Appli
37	80.5	5.3	1057	4	US-09-541-782-10	Sequence 10, Appli
38	80.5	5.3	1073	4	US-09-541-782-6	Sequence 6, Appli
39	80	5.3	472	1	US-08-496-631-2	Sequence 2, Appli
40	80	5.3	2930	4	US-09-417-822-2	Sequence 2, Appli
41	79.5	5.3	516	4	US-09-147-009-2	Sequence 2, Appli
42	79.5	5.3	835	4	US-09-291-839-2	Sequence 2, Appli
43	79	5.2	333	4	US-08-960-780-44	Sequence 44, Appli
44	79	5.2	333	4	US-09-073-898-44	Sequence 44, Appli
45	79	5.2	524	3	US-08-557-210A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-720-817-3

; Sequence 3, Application US/09720817

; Patent No. 6340579

; GENERAL INFORMATION:

; APPLICANT: Masai, Eiiji

; APPLICANT: Fukuda, Masao

; APPLICANT: Katayama, Yoshihiro

; APPLICANT: Nishikawa, Seiji

; APPLICANT: Hotta, Yasushi

; TITLE OF INVENTION: NOVEL GENE AND TRANSFORMANT CARRYING THE SAME

; FILE REFERENCE: 20096650XOPT

; CURRENT APPLICATION NUMBER: US/09/720.817

; CURRENT FILING DATE: 2001-06-04

; PRIOR APPLICATION NUMBER: JAPAN 10/203200

; PRIOR FILING DATE: 1998-07-17

; PRIOR APPLICATION NUMBER: PCT/JP99/03410

; PRIOR FILING DATE: 1999-06-25

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 315

; TYPE: PRT

; ORGANISM: Sphingomonas sp. CR-0310201

US-09-720-817-3

Query Match	8.3%	Score 124.5	DB 4	Length 315
Best Local Similarity	20.4%	Pred. No. 7.2e-05		
Matches	65	Conservative	55	Mismatches 147
				Indels 51
				Gaps 10
Qy	11	VVVGVRAGSRMRDLRNPSSAFNLNIGFVSR-----ELGSDIGVQO---ISLEDAL	63	
Db	3	IALAGAGAFGEKHLNIDG----VEIVSIIRKAEQAQAEVAAKYGAKHSGTDLSEAL	58	
Qy	64	SSQVEVAYICSESSSHEDYIRQFLNAGKHLVVEYPMTSLSLAAQELWELAEQKGLVHE	123	
Db	59	ARDDVDVILCTPTQMHAEQAIALACNAGKHVQVEIPLADSWADAEAVMKKSOETGLVCMV	118	
Qy	124	EHVELLMEEFAFLKEVVGKDL-----LKGS-----LLFTSDPLE	158	
Db	119	GTRFRNFPSHOYIHNKIVAGELAIQOMDVQTYFFRRKNMNAKGEPRSMTHLLMHAHAHT	178	
Qy	159	EDRFQFPAFGISRLTSLVSLFGELSLVSATLEERKEQYMKMTVCLETEKKSPLS----	214	
Db	179	VDLFAYQAGKIVQANAVOGPIHPELGIANDMSIQKSETGAICTLSLNFNDGPIGTFR	238	
Qy	215	WIEEKGPGLKRNRYLSHFHFKSGLE--NVPNVGVNKN--IFLKDQNIYFVKLLGQFSEKEL	271	
Db	239	YICDNGTWIARYDDL----VTGKEPVDVSKVDVSMNGIELQDREFIAAIREGPRENSSV	294	

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-980-357-19

Query Match 6.5%; Score 98; DB 3; Length 988;
Best Local Similarity 20.8%; Pred. No. 0.24;
Matches 66; Conservative 50; Mismatches 105; Indels 96; Gaps 14;
Qy 21 SVRMDLNP-----HPSSAFNLIGFVSRRELGSIDGVQOQISLSDALSSOQVEVAYICS 75
Db 55 AVQLAVLRPGWYTHIKSIDPSVQIYSKO-----IGVSSSL-DHPQRENTLWDHLK 108
Qy 76 ESSSHEDYIROFLNAGKHVLYEYPMTSLAAQELWELAEQKGVLYHEHVLLMEEF 135
Db 109 EIRSEYDEV-----TFTLSEYRMTF-----KYLHQLALENGDAIH-----LLHECIDF 151
Qy 136 LKKEVVGDKLKGSLFTSDPLEDRFGFPAPSGISRLTWLVSLFGELSLVSATLEERKE 195
Db 152 LRK-----NKKILPAITTLRMVWEARMAAEKKLFTVSKSLTN 190
Qy 196 DOYKMTVCLETE-----KKSPLSWIEKGGP-----LKNRYLSFHFKSGSLEN 240
Db 191 EKEKEGIIISQHPSESNTILGWKEP-PGHPSPETFLKTIERLEYI-----RGMDET 245
Qy 241 VNVGVNKNIFLK-----DQN-----IFVKLIGOFSEKELAAEKRR 277
Db 246 VQISHLRNRLQLSLGSRVEPYAFRDFQENKRYSLTIYLLQLTQELTKAFEIHDRQ 305
Qy 278 ILHCLG-----LAEETOK 290
Db 306 ILSLSKGRKAQOEIQK 322

RESULT 9

US-08-402-217A-2
; Sequence 2, Application US/08402217A
; Patent No. 5587301
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,217A
; FILING DATE: 10-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-002805
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-402-217A-2

Query Match 6.0%; Score 90.5; DB 1; Length 351;
Best Local Similarity 23.4%; Pred. No. 0.3;
Matches 68; Conservative 42; Mismatches 83; Indels 97; Gaps 14;
Qy 54 VQOISLE-DALSSQOEVEA-----YICS-----ESSSHEDYIROFLNAGKHVLYEYPM 100
Db 18 LQOKEQLIDSLLLQOKEKELSSSLHOKLCSFOEEMAKEKNLFEELKQTLD----- 66
Qy 101 TSLAAAEELWELAEQKGVLYHEHVLLMEEFALKKKEVVGK--DLLKGSLLFTS----- 154
Db 67 --ELDKLQOKEQEAERLYKQL--EEAKSRAEELKLEELKKEAELEKSSAAHTQATLL 123
Qy 155 -----DPLEEDRFPGFPAPSGISRLTWLVSLFGELSLVSATLEER----- 193
Db 134 LEEKYDSMVQSLDVTYAFESYKALT-----ASEIEDKLKLENSLOEKVAKAGKNAEDVQ 178
Qy 194 -----KEDQYKMTVCLETEKKSPLSWIEKGGPGLKRNRYLSFHFKSGSLENVPMVG 245
Db 179 HQILATESNQEYVRMLDLQOT--KSALKETEIK-----EITVSFLQKITDLOQO---- 226
Qy 246 VKNITFLDKQNTIFVQKLL-----GQFSEKELAAEKRRILHCLGLAEIEIOKY 291
Db 227 -----LKQEBDFRQKLEDEGRKAKEKENTTAE-----LTEEINKW 262

RESULT 10

US-08-700-178-2
; Sequence 2, Application US/08700178
; Patent No. 5783669
; Patent No. 5783669 5700912
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700.178
; FILING DATE: August 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,217
; FILING DATE: March 10, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0028-1 DIV
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-700-178-2


```
Query Match          5.8%; Score 88; DB 3; Length 551;
Best Local Similarity 22.1%; Pred. No. 1.1;
Matches 64; Conservative 42; Mismatches 110; Indels 74; Gaps 13;

Qy 1 MNAEPERKFGVVVGVGVRGRVSRMRDLRN-----PHPSAFLNLIGFVSRRELQ 49
Db 144 LEKQPKR---AVIVGAGFLEMAENLKRGLVNVVIEQAPH-----ILP 185

Qy 50 SIDGVOQISLEDAALSSQVEVAYTCSESSSHEDYIROFLNAGKHVLYEYPMTL-----102
Db 186 TLDEMAAFIEKELSHOGVEVI-----TSHA--VAGFEDHGKRLRLDDGRITPADLVIL 237

Qy 103 -----SLAAQELWELABQKVLHEEHVELLMEEFA-----FLKKEVVGVKDLKGS 149
Db 238 SIGVRPNQLAVTAGI-ELGTRGILVDERYQTNIPOIYAVGDALVVKQOITGKDAL----293

Qy 150 LLFTSDPLEEDRFGFPFSGISRL-----TWLSLFGELSLVSATLEERKEDQYMKMTV 203
Db 294 ISLASPANROQRQVADTTISGISRNOGGIGTAIIRTFG-MTAASGTLSERTAKENELSF 352

Qy 204 CLETEKKSPLSWIEKGCPLRNRYLSPHFKSGSLENVPNV---GVNKNI 250
Db 353 VIHVGKDHASYYPEATDILLK---LIFHPETGEIYGAQGVGAKGVDRRI 399

RESULT 13
US-07-857-224B-9
; Sequence 9, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264.
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: bovine
; FEATURE: protein kinase; Table 8 Column 9
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.
; AUTHORS: Quinn, A. M.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
```

```
; PAGES: 42-52
; DATE: 1988
US-07-857-224B-9

Query Match          5.8%; Score 87; DB 2; Length 264;
Best Local Similarity 21.8%; Pred. No. 0.45;
Matches 39; Conservative 33; Mismatches 73; Indels 34; Gaps 10;

Qy 9 FGVV-VVGVGRAGSVRMRLRNPSPSSAFLNLIGFVSRRELGSIDGVOQISLEDAALSSQE 67
Db 3 FNIIDTLGVGGFGRVQLKSESKTFAMKIL---KKRHI--VDTROQ---EHIRSEKQ 54

Qy 68 VEVAYTCSESSSHEDYIROFLNAGKHVLYEYPMTLSLAAQELWELABQKVLHEE---124
Db 55 I-----MQAHSDFIVRLYRTFKDSKLY-MLMEACLGELGELWTILDRGSGFEDSTTF 106

Qy 125 HVELLMEEFAEL-KKEVVGVKDLKGSLLFTSDPLEEDR-----FGPPAFSGISRLTW 175
Db 107 YTACVVEAFAYLHSGIIIRDLKPNLIL-----DHRGYAKLVDFGFPAKKIGFGKLTW 159

RESULT 14
US-08-676-967-1
; Sequence 1, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 759 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-08-676-967-1

Query Match          5.7%; Score 86; DB 1; Length 759;
Best Local Similarity 22.3%; Pred. No. 2.9;
Matches 58; Conservative 43; Mismatches 99; Indels 60; Gaps 11;

Qy 22 VRMRDLRNPSPSSAFLNLIGFVSRRELGSIDGVOQISLEDAALSSQVEVAYTCSESSSHE 81
Db 267 IQKRAVKRPAPAKSSDSEEDSDLEESDIDGELAQSDTSEQEDKAVOVSNKKRRK 326

Qy 82 DYIROFLNAGKHVLYEYPMTLISL-AAQELWELABQKQK-----VLHEEHVELLMEEFA 134
Db 327 --LPSDVNKGTVFIR---NLSFDSEEEELGELLQOFGELKYLVRIVLHPDTEHSGKCAFA 381
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:40:53 ; Search time 67.0997 Seconds
(without alignments)
489.985 Million cell updates/sec

Title: US-09-606-129A-1

Perfect score: 1508

Sequence: 1 MNAEPERKFGVVVGVGRG.....RILHCLGLABEIQKYCCSRK 296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225	14.9	60	21	AAG02024 Human secreted pro
2	186	12.3	63	21	AAG02023 Human secreted pro
3	152.5	10.1	434	22	ABG21584 Novel human diagno
4	148	9.8	241	22	ABG21582 Novel human diagno
5	148	9.8	389	22	ABG29497 Novel human diagno
6	129	8.6	335	22	ABG60024 Drosophila melanog
7	124.5	8.3	315	21	AAU68460 Sphingomonas sp. C
8	122.5	8.1	226	22	AAU49252 Propionibacterium
9	122.5	8.1	280	21	AAU68459 Sphingomonas sp. C
10	121.5	8.1	175	18	AAW28328 Amino acid sequenc
11	120.5	8.0	337	22	AAG92054 C glutamicum prote

12	120.5	8.0	337	22	AAB79386 Corynebacterium gl
13	119.5	7.9	346	22	AAU36865 Staphylococcus aur
14	119	7.9	341	22	AAU36865 Putative P. abyssal
15	116.5	7.7	341	22	AAU34111 Staphylococcus aur
16	110	7.3	357	22	AAU34110 Staphylococcus aur
17	110	7.3	359	22	AAU36864 Staphylococcus aur
18	108	7.2	229	19	AAU77759 Glucose-fructose o
19	108	7.2	329	22	AAE00480 Streptomyces galli
20	108	7.2	344	22	AAU04826 Micromonospora eve
21	107	7.1	332	22	AAU39310 SpnN protein invol
22	107	7.1	336	22	AAU70956 S. spinosa protein
23	106.5	7.1	315	22	AAU96316 Putative P. abyssal
24	106.5	7.1	344	22	AAU43141 Propionibacterium
25	105.5	7.0	307	22	AAU98892 E. coli growth and
26	105	7.0	327	22	AAU03634 Group B Streptococ
27	102	6.8	753	22	ABG20518 Novel human diagno
28	100.5	6.7	1316	22	AAU491126 Mycobacterium tube
29	98.5	6.5	702	22	AAU49114 Propionibacterium
30	97.5	6.5	413	22	AAU56511 Propionibacterium
31	97	6.4	219	19	AAU98401 H. pylori GHP0 138
32	97	6.4	988	13	Transposase from g
33	96	6.4	424	21	AAU43883 Human cancer assoc
34	96	6.4	1374	22	AAU69070 Human male enhance
35	95	6.3	359	22	AAU93130 C glutamicum prote
36	95	6.3	359	22	AAU79395 Corynebacterium gl
37	94	6.2	417	22	ABG67972 Drosophila melanog
38	93	6.2	623	22	AAU37906 Streptococcus pneu
39	92.5	6.1	867	22	AAU96297 Putative P. abyssal
40	91	6.0	337	22	AAU89930 C glutamicum prote
41	90.5	6.0	237	22	ABG24349 Novel human diagno
42	90.5	6.0	341	22	AAU89934 C glutamicum prote
43	90.5	6.0	341	22	AAU79394 Corynebacterium gl
44	90.5	6.0	351	17	AAU01052 Human umbilical ve
45	90.5	6.0	433	22	AAU4397 Staphylococcus aur

ALIGNMENTS

RESULT 1

AAG02024
ID AAG02024 standard; Protein: 60 AA.

XX AAG02024;

XX AC AAG02024;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein, SEQ ID NO: 6105.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX KW gene therapy; chromosome mapping.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX XX (GSEST) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI: 2000-500381/45.

XX XX N-PSDB; AAC02030.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 13; SEQ ID 6105; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number

CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs

CC were prepared from total human RNAs or polyA+ RNAs derived from 30

CC different tissues. EST sequences usually correspond mainly to the 3'

CC untranslated region (UTR) of the mRNA because they are often obtained

CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for

CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in

CC those cases where longer cDNA sequences have been obtained, the full 5'

CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'

CC ends and can therefore be used to obtain full length cDNAs and genomic

CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and

CC chromosome mapping procedures. They are used to obtain upstream

CC regulatory sequences and to design expression and secretion vectors.

XX

SQ Sequence 60 AA;

Query Match 14.9%; Score 225; DB 21; Length 60;

Best Local Similarity 97.9%; Pred. No. 2.2e-14;

Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 MRDLRNPSPSAFLNLIGFVSRRELGSIDGVQQTSLSDALSQVEV 70

|||||

DB 1 MRDLRNPSPSAFLNLXGFSRRELGSIDGVQQTSLSDALSQVEV 47

RESULT 2

AG02023

ID AAG02023 standard; Protein: 63 AA.

XX

AC AAG02023;

DT 06-OCT-2000 (first entry)

XX

DE Human secreted protein, SEQ ID NO: 6104.

XX

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping.

KW Homo sapiens.

OS

XX EP1033401-A2.

PN

XX 06-SEP-2000.

PD

XX 21-FEB-2000; 2000EP-0200610.

PF

XX 26-FEB-1999; 99US-0122487.

PR

XX (GEST) GENSET.

PA

XX Dumas Milne Edwards J, Duclert A, Giordano J;

PI

XX WPI; 2000-500381/45.

DR

DR N-PSDB; AAC02029.

XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX

PS Claim 13; SEQ ID 6104; 71pp + CD-ROM; English.

XX

CC The present sequence is a polypeptide encoded by one of a large number

CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs

CC were prepared from total human RNAs or polyA+ RNAs derived from 30

CC different tissues. EST sequences usually correspond mainly to the 3'

CC untranslated region (UTR) of the mRNA because they are often obtained

CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for

CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in

CC those cases where longer cDNA sequences have been obtained, the full 5'

CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'

CC ends and can therefore be used to obtain full length cDNAs and genomic

CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and

CC chromosome mapping procedures. They are used to obtain upstream

CC regulatory sequences and to design expression and secretion vectors.

XX

SQ Sequence 63 AA;

Query Match 12.3%; Score 186; DB 21; Length 63;

Best Local Similarity 94.9%; Pred. No. 1.4e-10;

Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNAEPERKFGVVVGVGRAGSVRMRDLRNPSPSAFLNL 39

|||||

DB 1 MNAEPERKFGVVVGVGRAGSVRMRDLRNPSPSAFLNL 39

RESULT 3

ABG21584

ID ABG21584 standard; Protein: 434 AA.

XX

AC ABG21584;

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #21575.

XX

Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

KW Homo sapiens.

OS

XX WC200175067-A2.

PN

XX 11-OCT-2001.

PD

XX 30-MAR-2001; 2001WO-US08631.

PF

XX 31-MAR-2000; 2000US-0540217.

PR

XX 23-AUG-2000; 2000US-0649167.

PR

XX (HYSE-) HYSEQ INC.

PA

XX Drmanac RT, Liu C, Tang YT;

PI

XX WPI; 2001-639362/73.

DR

DR N-PSDB; AAS85771.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX

PS Claim 20; SEQ ID NO 51943; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABC30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 389 AA;

Query Match 9.8%; Score 148; DB 22; Length 389;
Best Local Similarity 29.9%; Pred. No. 8.3e-06;
Matches 43; Conservative 28; Mismatches 53; Indels 20; Gaps 4;

QY 59 LEDALSSQEVAVYTCSSSSSHEDYIROFLNAGKHVLVEYPMTLSLAAQELWELAEQKG 118

Db 222 LDEVNDPDKLVVYVCHADSHFEYAKRALEAGNVLVEKPTPTLAAQKELFALAKSGK 281

QY 119 KVLHEERVELLMEEFAFLKKEVWGKDLKGSLLFTSD-----PLEEDRFQFP---AFSG 169

Db 282 LTVTPYQNRFRDSCFLTAKKAIESGKL--GEIVEVESHFYIRPVAETKPGLPQDGFYFG 339

QY 170 ISRL-----TWLVSLFCELS 184

Db 340 LGALFTNQGGFFKSLLSIFADSS 363

RESULT 6

ABB60024

ID ABB60024 standard; Protein; 335 AA.

XX ABB60024;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 6864.

XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL04127.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

XX Disclosure; SEQ ID NO 6864; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 335 AA;

Query Match 8.6%; Score 129; DB 22; Length 335;
Best Local Similarity 27.7%; Pred. No. 0.00047;
Matches 44; Conservative 28; Mismatches 62; Indels 25; Gaps 6;

QY 1 MNABPERKFGVVVVG-----VGRAGSVYMRDLRNPSPSSAFNLNIG-----FYSRREL- 48

Db 1 MQSOANLWNGIAAAGRITQDFVTALGTVE----KSRHVVAADVDDGQRAQOFAQRNQIP 56

QY 49 GSIDGVOQISLEDALSSQEVAVYTCSSSSSHEDYIROFLNAGKHVLVEYPMTLSLAAQ 108

Db 57 RHYGFDALAD-----REVEVYVGTLPNPFYAVVHMLARGKNVICETPMCLSVQAK 111

QY 109 ELWELAEQKGKVLHEERVELLMEEF-----AFLKKEVWGK 143

Db 112 ELYTLAEQRGVFLMEGMWSRFFPSYDRURELLKNDVIGE 150

RESULT 7

AAAY68460

ID AAY68460 standard; Protein; 315 AA.

XX AAY68460;

XX 20-APR-2000 (first entry)

XX Spingomonas sp. CR-0310201 CHMSD SEQ ID NO:3.

XX Alpha-hydroxy-gamma-carboxymuconic acid-eta-semialdehyde dehydrogenase;
XX 2-pyrone-4,6-dicarboxylic acid; 3,4-dihydroxybenzoic acid.

XX Spingomonas sp.

XX WO200004134-A1.

XX 27-JAN-2000.

XX 25-JUN-1999; 99WO-JP03410.

XX 17-JUL-1998; 98JP-0203200.

XX (MAZN) COSMO RES INST.

XX (MAZN) COSMO OIL CO LTD.

XX Masai E, Fukuda M, Katayama Y, Nishikawa S, Hotta Y;

XX WPI; 2000-147784/13.

XX N-PSDB; AAZ88148, AAZ88151.

XX New alpha-hydroxy-gamma-carboxymuconic acid eta-semialdehyde
XX dehydrogenase for industrial production of 2-pyrone-4,6-dicarboxylic
XX acid -

XX Claim 1; Page 26-27; 34pp; Japanese.

XX The present sequence represents an alpha-hydroxy-gamma-carboxymuconic
XX acid-eta-semialdehyde dehydrogenase (CHMSD) originating in
XX Spingomonas sp. CR-0310201. CHMSD can be used in the production of
XX 2-pyrone-4,6-dicarboxylic acid by culture of a transformant comprising
XX CHMSD in the presence of a suitable precursor such as
XX 3,4-dihydroxybenzoic acid (which is converted in situ into
XX alpha-hydroxy-gamma-carboxymuconic acid-eta-semialdehyde).

```
SQ Sequence 315 AA;
Query Match 8.3%; Score 124.5; DB 21; Length 315;
Best Local Similarity 20.4%; Pred. No. 0.0012;
Matches 65; Conservative 55; Mismatches 147; Indels 51; Gaps 10;

Qy 11 VVVGVRAGSVRMRLRNPHPSAFLNLIGFVSR-----ELGSDGVQV--ISLEDA 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 IALAGAGAFGEKHLGDKNIDG---VEIVSTIIRKAEQAEEAAKYGAKHSGTDLSEAL 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 64 SSQEVEVAYICSESSSHEDYIROFLNAGKHVLYVEYPTWLSLAAAGELWELAEQKGVLHE 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 ARDDVDVILCTPTQMAEQIAACNAGKHVQVEIPLADSWADAEAVNMKKSQETGLVCVM 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 124 EHVLELMEFAFLKKEVVGKDL-----LKGS-----LLFTSDPLE 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 GHTRRPNSHOYIHNKIVAGELAIQMDVQYFFRRKNNAKGEPRSWTDHLLHHAHT 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 159 EDRGFPAPFSGISRLTWLVSLFGLSLVSATLEERKEDQYKMWVCLTEKKSPLS----- 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 VDLFAYQAGKIVQANAVQGPVHPHGLGAMDMNSIQLKSETGAICTLSLSFNNDGPIGTFFR 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 215 WIEKGGCLKKNRYLSFRFKSGSLE--NVPNVGVNKN-IFLKDQNIQVQKLLGQFSEKEL 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 YICDNGTWIARYDDL-----VTGKEEPVDVSKVDVSMNGIELQDREFFIAIREGPREPNSSV 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 272 AAERKRLHLCIGLAEIEIQ 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 A----RVLDYRVLGELE 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
AAU49252
ID AAU49252 standard; Protein: 226 AA.
XX AC AAU49252;
XX DT 27-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein #10148.
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KW dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN W0200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US12865.
XX PR 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX SK Skey YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX DR WPI: 2001-616774/71.
XX DR N-PSDB; AAS59545.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris.
XX FS Claim 6; SEQ ID No 10447; 1069pp; English.
XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
```

```
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 226 AA;
Query Match 8.1%; Score 122.5; DB 22; Length 226;
Best Local Similarity 23.9%; Pred. No. 0.0012;
Matches 48; Conservative 21; Mismatches 79; Indels 53; Gaps 4;

Qy 13 VVGVRAGSVRMRLRNPHPSAFLNLIGFVSRRELSTGVOQISLEDAISSQBEVEA- 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 23 MVGIGSGCKNHVRNL-----AIDGVDLVAIADASKDPFGVAG 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 72 -----YICSESSSHEDYIROFLNAGKHVLYVEYPTWLSLAAAGELWE 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 DLPVLPDQVAVTETGIDYCVVAAPTKEEIGLKLAEAGVHALIEKPLAYDTAAATRLAE 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 113 LAEQKGVLHEHVLELMEFAFLKKEVVGKDLKGLLFTSDPLEEDRFPGFPGISLR 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 AFESKGLGVAGVHTERFNPALOSLRKRLGDL--GDLYQVATRRQ-----GHSR 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 173 LTWLVSFLGELSLVSATLEER 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 PVSLSALSRLSLATLTSR 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
AAU68459
ID AAU68459 standard; Protein: 280 AA.
XX AC AAU68459;
XX DT 20-APR-2000 (first entry)
XX DE Sphingomonas sp. CR-0310201 CHMSD SEQ ID NO:1.
XX KW Alpha-hydroxy-gamma-carboxymuconic acid-eta-semialdehyde dehydrogenase;
XX KW 2-pyrone-4,6-dicarboxylic acid; 3,4-dihydroxybenzoic acid.
XX OS Sphingomonas sp.
XX PN W02000004134-A1.
XX PD 27-JAN-2000.
XX PF 25-JUN-1999; 99WO-JP03410.
XX PR 17-JUL-1998; 98JP-0203200.
XX PA (MAZN ) COSMO RES INST.
XX PA (MAZN ) COSMO OIL CO LTD.
XX PI Masai E, Fukuda M, Katayama Y, Nishikawa S, Hotta Y;
XX DR WPI; 2000-147784/13.
XX DR N-PSDB; AAZ88147.
```

```

XX  New alpha-hydroxy-gamma-carboxymuconic acid eta-semialdehyde
PT  dehydrogenase for industrial production of 2-pyrone-4,6-dicarboxylic
PT  acid
XX
XX  Claim 1; Page 23-25; 34pp; Japanese.
XX
XX  The present sequence represents an alpha-hydroxy-gamma-carboxymuconic
CC  acid-eta-semialdehyde dehydrogenase (CHMSD) originating in
CC  Shingomonas sp. CR-0310201. CHMSD can be used in the production of
CC  2-pyrone-4,6-dicarboxylic acid by culture of a transformant comprising
CC  CHMSD in the presence of a suitable precursor such as
CC  3,4-dihydroxybenzoic acid (which is converted in situ into
CC  alpha-hydroxy-gamma-carboxymuconic acid-eta-semialdehyde).
XX
SQ  Sequence 280 AA;
Query Match 8.1%; Score 122.5; DB 21; Length 280;
Best Local Similarity 22.5%; Pred. No. 0.0016;
Matches 32; Conservative 33; Mismatches 66; Indels 11; Gaps 3;
QY 11 VVVVGVGRAGSVRMRDLRNPSPSSAFNLNIGFVSR-----ELGSDGVQV--ISLEDAL 63
DB 3 IALAGAGAFGEKHLGDKNIDG----VEIVSIIRKAEQAEEVAAYKAGKHSHTDLSEAL 58
QY 64 SSQVEVAYICSSSSSHEDYIROFLNAGKHVLYEYPMTLSLAAQELWELAEQKGVILHE 123
DB 59 ARDDVDVAVILCTPQMHAEQAIAICMNAKGKHVQVEIPLADSWADAEAVMKKSQETGLVCMV 118
QY 124 EHVLLMEEFAPLKKVGVGKDL 145
DB 119 GHTRRFNPSHOYIHNKIVAGEL 140
RESULT 10
AAW28328
ID AAW28328 standard; Protein: 175 AA.
XX
XX AAW28328;
XX
XX 01-SEP-1998 (first entry)
XX
XX Amino acid sequence of glucose-fructose oxidoreductase precursor.
XX
XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW Staphylococcal gene; regulatory element; bacterial gene expression;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW toxic shock syndrome.
XX
XX Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FH Misc-difference 1..175
FT /note= "residues designated X are not defined in
FT the specification"
XX
XX WO9730070-A1.
XX
XX 21-AUG-1997.
XX
XX 19-FEB-1997; 97WO-US02318.
XX
XX 20-FEB-1996; 96US-0011888.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX
XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
XX
XX WPI: 1997-424969/39.
XX
XX N-PSDB; AAT84223.
XX
XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used

```

```

PT to isolate antimicrobial compounds, and in vaccines against S.
PT aureus infection
XX
XX Claim 6; Page 610; 989pp; English.
XX
XX The present sequence represents a Staphylococcus aureus protein, that,
CC based on homology with a zymomonas sp. protein, is believed to be a
CC glucose-fructose oxidoreductase precursor. The DNA sequence was isolated
CC from a library of clones of S. aureus WCUH 29 in Escherichia coli. The
CC DNA sequence can be used in the construction of ribozymes and antisense
CC sequences to control the expression of Staphylococcal genes. The DNA
CC sequence is also useful as a source of regulatory elements for the
CC control of bacterial gene expression. The present protein may be used
CC to produce vaccines to enable a host to produce specific antibodies
CC with antibacterial action. These vaccines and antibodies would protect
CC a host against invasion by S. aureus, and conditions relating to
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC skin syndrome, and toxic shock syndrome.
XX
SQ Sequence 175 AA;
Query Match 8.1%; Score 121.5; DB 18; Length 175;
Best Local Similarity 27.9%; Pred. No. 0.001;
Matches 43; Conservative 22; Mismatches 66; Indels 23; Gaps 4;
QY 8 KFGVVGVGRAGSVRMRDLRNPSPSSAFNLNIGFVSRRELGSIDGVQVQISLEDAISSQE 67
DB 3 KLVGVGVG--GLAQRHI-----PALLKLDVSLVAVQDINTVQMIDVAKRFNIPQ 54
QY 68 -----VEVAYICSSSSSHEDYIROFLNAGKHVLYEYPMTLSLAAQELWELAEQ 116
DB 55 AVETPSELFKLVDVAVICTPNKPHADLSIEALNHGVHVLCEKPMXMTTECDRMIEAANK 114
QY 117 KGKVL----HEEHVELLMEEFAPLKKVGVGKDLL 146
DB 115 NKKLLTVAYHYRHTDVAMTAKKAIEAGVVGKPLV 148
RESULT 11
AAG92054
ID AAG92054 standard; Protein: 337 AA.
XX
XX AAG92054;
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 5808.
DE
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR
XX 07-APR-2000; 2000JP-0159162.
PR
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
XX
XX N-PSDB; AAH67273.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing

```

expression profile or pattern of a gene and identifying homologous gene

Claim 17; SEQ ID NO: 5808; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Corynebacterium glutamicum glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Corynebacterium, and identifying a homologue of a gene derived from corynebacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Query Match 8.0%; Score 120.5; DB 22; Length 337;
Best Local Similarity 25.1%; Pred. No. 0.0032;
Matches 42; Conservative 25; Mismatches 69; Indels 31; Gaps 4;

QY 11 VVVGVGRAGSVRMRLRN-----PHPSAFNLNLGIFVSRRELGSIDGVQOIS 58
DB 5 LALIGAGRIGSNHARLITNHVIGSELVAVDPTPNAETL-----ADELGA---VAFSN 54
QY 59 LEDALSSQEEVAYICSSSHEDYIROFLNAGKHLVVEYPMFTLSLAAQELWELAEQKG 118
DB 55 PDDVLTRODDIDAVLIATPARTHADLVVKAAGKHVFEKPMAVTLEDADRINAAREAN 114
QY 119 KYLHEEHVELLMEEFAFLKKEVVGKDLKGSLL-----FTSDP 156
DB 115 TVLVQGFNRRFAAGFAAARARIDAGDIGTPTQLLRSVTRDPGPFADP 161

RESULT 12
AAB79386
ID AAB79386 standard; Protein; 337 AA.
AC AAB79386;
XX 30-APR-2001 (first entry)
XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:288.
XX Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carboxylate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

Corynebacterium glutamicum.

WO200100844-A2.

04-JAN-2001.

23-JUN-2000; 2000WO-IB00943.

25-JUN-1999; 99US-0141031.

08-JUL-1999; 99DE-1031412.

08-JUL-1999; 99DE-1031413.

08-JUL-1999; 99DE-1031419.

08-JUL-1999; 99DE-1031420.

08-JUL-1999; 99DE-1031424.

08-JUL-1999; 99DE-1031428.

08-JUL-1999; 99DE-1031431.

08-JUL-1999; 99DE-1031433.

PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX (BADI) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

WPI; 2001-061975/07.

N-PSDB; AAF71503.

New isolated Corynebacterium glutamicum nucleic acid encoding a sugar metabolism and oxidative phosphorylation protein for production or modulation of production of fine chemicals e.g. amino acids, carbohydrates or enzymes.

Claim 20; Page 550-551; 1246pp; English.

AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243 to AAB 79633 which are involved in carbon metabolism and energy production. The C. glutamicum SMP gene can be used in vectors (ii) for expression in host cells and production or modulation of production of fine chemicals, such as, an organic acid, a proteinogenic or nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a polyketide, or an enzyme. The presence of (i) or SMP proteins (iii) encoded by them are used for diagnosing the presence or activity of Corynebacterium diphtheriae in a subject. (i), (ii), (iii) or host cells containing them are used to map genomes of organisms related to C. glutamicum, identify and localise C. glutamicum sequences of interest, in evolutionary studies, in determining SMP protein regions required for function, in modulating SMP protein activity, in modulating the metabolism of sugars, and in modulating high-energy molecule production in a cell (i.e. ATP, NADPH).

Sequence 337 AA;

Query Match 8.0%; Score 120.5; DB 22; Length 337;
Best Local Similarity 25.1%; Pred. No. 0.0032;
Matches 42; Conservative 25; Mismatches 69; Indels 31; Gaps 4;

QY 11 VVVGVGRAGSVRMRLRN-----PHPSAFNLNLGIFVSRRELGSIDGVQOIS 58

DB 5 LALIGAGRIGSNHARLITNHVIGSELVAVDPTPNAETL-----ADELGA---VAFSN 54

QY 59 LEDALSSQEEVAYICSSSHEDYIROFLNAGKHLVVEYPMFTLSLAAQELWELAEQKG 118

DB 55 PDDVLTRODDIDAVLIATPARTHADLVVKAAGKHVFEKPMAVTLEDADRINAAREAN 114

QY 119 KYLHEEHVELLMEEFAFLKKEVVGKDLKGSLL-----FTSDP 156

DB 115 TVLVQGFNRRFAAGFAAARARIDAGDIGTPTQLLRSVTRDPGPFADP 161


```
Qy 170 I-----SRLTWL-----VSLFGLSVSATLEERKEDQYMKMTVCLETEKK--SPL 213
      :          :          :          :          :          :
Db 174 VIVDLGVHVTDFLRWFESPIEVFAVCKAIRGEARVKNKTHDHVVMFIKFEKGKTIGEV 233
      :          :          :          :          :          :
Qy 214 SWIEKGGGLKRNRYLSHF-----KSGSLE----NVPNGVKNKNIFLKQONIPVQKLL 263
      :          :          :          :          :          :
Db 234 SWSVPMP-----KYGVFYHLDIIGKNGRIRYTPMDTPVVGKSSFEMPR---FSPML 285
      :          :          :          :          :          :
Qy 264 GQSEKELAAEKKRILHCL 282
      :          :          :          :          :          :
Db 286 STFPQ-AFEALRHFEFCI 303
      :          :          :          :          :          :

RESULT 15
AAU34111
ID AAU34111 standard; Protein; 341 AA.
AC AAU34111;
XX
DT 13-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #387.
XX
KW Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RI, Xu HH;
XX
WPI; 2001-611495/70.
DR N-PSDB; AAS51970.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 5607; 51lpp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
```

```
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 341 AA;

Query Match 7.7%; Score 116.5; DB 22; Length 341;
Best Local Similarity 27.8%; Pred. No. 0.0078;
Matches 42; Conservative 21; Mismatches 65; Indels 23; Gaps 4;

Qy 11 VVVVGVRAGSVVRMDLRNPHPSAFLNLIGFVSRRELGSIDGVOQISLEDALESSOE--- 67
      ||||| : : : : : || : : : : : : : : : : : : : : : : : : : : :
Db 3 VGVIGVG--GIAQDRHI-----PALLKLDKTVSLVAVQDINTVOMIDVAKRFNIPHYVE 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 68 -----VEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAAQELWELAEQK 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 TPSELFKLVDVAVICTPNKFHADLSIEALNHGVHVLCEKPMAMTTECDRMIEAANKNCK 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 120 VL-----HEEHVELLMEEFAFLKKEVVGKDLL 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 LLTVAYHYRHTDVAAMTAKKAIESGVVGKPLV 145

Search completed: November 1, 2002, 14:47:03
Job time : 70.0997 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:08 ; Search time 53.7749 Seconds
(without alignments)
952.238 Million cell updates/sec

Title: US-09-606-129A-3

Perfect score: 1508

Sequence: 1 MNTEPERKFGVVVGVGRAG.....RILHCLGLAEIQQYCCSRK 296

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL_19:*
- 2: sp_archaea:*
- 3: sp_bacteria:*
- 4: sp_fungi:*
- 5: sp_human:*
- 6: sp_invertebrate:*
- 7: sp_mammal:*
- 8: sp_mhc:*
- 9: sp_organelle:*
- 10: sp_phase:*
- 11: sp_plant:*
- 12: sp_rodent:*
- 13: sp_virus:*
- 14: sp_vertebrate:*
- 15: sp_unclassified:*
- 16: sp_rvirus:*
- 17: sp_bacteriap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1503	99.7	296	4 Q96QL4	Q96ql4 homo sapien
2	1502	99.6	296	4 Q9BRW8	Q9brw8 homo sapien
3	1264.5	83.9	295	11 Q9CV64	Q9cv64 mus musculus
4	1264.5	83.9	303	11 Q9DD21	Q9dd21 mus musculus
5	463	30.7	92	4 Q95019	Q95019 homo sapien
6	217	14.4	328	16 P72782	P72782 synechocyst
7	157.5	10.4	371	17 Q97UH0	Q97uh0 sulfolobus
8	143	9.5	393	16 P94437	P94437 bacillus su
9	141	9.4	318	17 O26961	O26961 methanother
10	137.5	9.1	382	16 Q986T3	Q986t3 rhizobium l
11	136.5	9.1	364	10 O23580	O23580 arabidopsis
12	132	8.8	360	16 Q9WZ54	Q9wz54 thermotoga
13	130.5	8.7	335	5 Q9VQB3	Q9vqb3 drosophila
14	130.5	8.7	368	3 O42896	O42896 schizosacch
15	129.5	8.6	324	16 Q9CFC7	Q9cfc7 lactococcus
16	129	8.6	319	2 Q93PS4	Q93ps4 comamonas t

17	127	8.4	351	16 Q98DB4	Q98db4 rhizobium l
18	124.5	8.3	369	3 O13991	O13991 schizosacch
19	123.5	8.2	315	2 Q9KWL3	Q9kwl3 pseudomonas
20	123.5	8.2	371	17 O59539	O59539 pyrococcus
21	123	8.2	330	2 O52566	O52566 amycolatops
22	122	8.1	338	16 Q92ET3	Q92et3 listeria in
23	121.5	8.1	323	16 Q92F05	Q92f05 listeria in
24	121.5	8.1	346	16 Q99X04	Q99x04 staphylococ
25	121	8.0	341	17 Q9UXX4	Q9uxx4 pyrococcus
26	120.5	8.0	369	2 Q9L3U7	Q9l3u7 streptomyc
27	120	8.0	346	2 Q54531	Q54531 streptomyc
28	119	7.9	308	16 Q92QY5	Q92qy5 rhizobium m
29	119	7.9	343	16 Q926S4	Q926s4 listeria in
30	117.5	7.8	357	16 Q9CIU1	Q9ciul lactococcus
31	116	7.7	328	2 Q9LCW9	Q9lcw9 streptomyc
32	116	7.7	358	2 O52495	O52495 listeria mo
33	115	7.7	378	2 Q9RDH5	Q9rdh5 streptomyc
34	115	7.6	358	16 Q929K9	Q929k9 listeria in
35	114	7.6	360	16 Q9HU35	Q9hu35 pseudomonas
36	114	7.6	371	2 Q9RK99	Q9rk99 streptomyc
37	112	7.4	332	2 Q9ALN5	Q9aln5 saccharopol
38	112	7.4	333	11 Q9DBB8	Q9dbb8 mus musculu
39	112	7.4	359	16 Q932K4	Q932k4 staphylococ
40	111.5	7.4	315	17 Q9UZ13	Q9uz13 pyrococcus
41	111.5	7.4	331	16 Q98CV0	Q98cv0 rhizobium l
42	111.5	7.4	349	16 Q929L3	Q929l3 listeria in
43	111	7.4	334	16 Q9WYP5	Q9wyp5 thermotoga
44	110.5	7.3	305	16 Q9CEL4	Q9cel4 lactococcus
45	110	7.3	314	2 Q93N64	Q93n64 coxiella bu

ALIGNMENTS

RESULT 1

ID	Q96QL4	PRELIMINARY;	PRT;	296 AA.
AC	Q96QL4:			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	UNKNOWN (PROTEIN FOR MGC:14706).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RC	TISSUE=PROSTATE;			
RA	Strausberg R.;			
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC008456; AAH08456.1; -			
SQ	SEQUENCE 296 AA; 33428 MW; 2CF2AA7FICDDB707 CRC64;			

Query Match	99.7%;	Score 1503;	DB 4;	Length 296;
Best Local Similarity	99.7%;	Pred. No. 1.1e-118;		
Matches 295;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MNTEPERKFGVVVGVGRAGSVRMRDLRNPSPSSAFNLNIGFVSRRRLGSDIGVQOQ1SLE	60	
Db	1	MNAEPERKFGVVVGVGRAGSVRMRDLRNPSPSSAFNLNIGFVSRRRLGSDIGVQOQ1SLE	60	
Qy	61	DALSSQEVAYICSESSSHEDYIROFLNAGKHVLPVMTLSLAAQELWELAEQKGV	120	
Db	61	DALSSQEVAYICSESSSHEDYIROFLNAGKHVLPVMTLSLAAQELWELAEQKGV	120	
Qy	121	LHEHVELLMEEFAFLKKEVGVKDLKGLLFTAGPLPEERFGPPAFSGISRLLTWLSLF	180	
Db	121	LHEHVELLMEEFAFLKKEVGVKDLKGLLFTAGPLPEERFGPPAFSGISRLLTWLSLF	180	
Qy	181	GELSIVATLEERKEDQYMKMTVCLETEKKSPLSWIEEKGPCLKRNRYLSFHFHSGSLEN	240	
Db	181	GELSIVATLEERKEDQYMKMTVCLETEKKSPLSWIEEKGPCLKRNRYLSFHFHSGSLEN	240	

```
QY 241 VPNVGNKNIFLKDQNTFVQKLLGFSEKELAAEKKRILHCLGLAEIQRKCCSRK 296
|||||
DB 241 VPNVGNKNIFLKDQNTFVQKLLGFSEKELAAEKKRILHCLGLAEIQRKCCSRK 296

RESULT 2
Q9BRW8 ID Q9BRW8 PRELIMINARY; PRT; 296 AA.
AC Q9BRW8; 2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE SMILAR TO BILIVERDIN REDUCTASE A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN, PRIMITIVE NEUROECTODERMAL;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005902; AA05902.1; -.
DR InterPro; IPR000683; GFO_IDH_Moca.
DR Pfam; PF01408; GFO_IDH_Moca; 1.
SQ SEQUENCE 296 AA; 33432 MW; A389AFDDB6ADBABF CRC64;

Query Match 99.6%; Score 1502; DB 4; Length 296;
Best Local Similarity 99.7%; Pred. No. 1.3e-118;
Matches 295; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNTPEPKFGVVVGVGRAGSVRMRLRNPHPSAFNLNLIGFVSRRRLGSDGVQOQISLE 60
|||||
DB 1 MNTPEPKFGVVVGVGRAGSVRMRLRNPHPSAFNLNLIGFVSRRRLGSDGVQOQISLE 60

QY 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120
|||||
DB 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120

QY 121 LHEHVLLMEEFALFKKVVGVKDLLKGLSLFTAGPLEEERFGPPAFSGISRLTWLSLF 180
|||||
DB 121 SHEHVLLMEEFALFKKVVGVKDLLKGLSLFTAGPLEEERFGPPAFSGISRLTWLSLF 180

QY 181 GELSLVSATLEERKEDQYMKMTVCLETEKKSPLSWIEKGPGLKRNRYLSFHFKSGSLEN 240
|||||
DB 181 GELSLVSATLEERKEDQYMKMTVCLETEKKSPLSWIEKGPGLKRNRYLSFHFKSGSLEN 240

QY 241 VPNVGNKNIFLKDQNTFVQKLLGFSEKELAAEKKRILHCLGLAEIQRKCCSRK 296
|||||
DB 241 VPNVGNKNIFLKDQNTFVQKLLGFSEKELAAEKKRILHCLGLAEIQRKCCSRK 296
```

```
RESULT 3
Q9CY64 ID Q9CY64 PRELIMINARY; PRT; 295 AA.
AC Q9CY64;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 2500001N03RIK PROTEIN.
GN 2500001N03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
```

```
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata K., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010847; BAB27219.1; -.
DR MGD; MGI:1917355; 2500001N03RIK.
DR InterPro; IPR000683; GFO_IDH_Moca.
DR Pfam; PF01408; GFO_IDH_Moca; 1.
SQ SEQUENCE 295 AA; 33524 MW; F2E1682BD77032A4 CRC64;

Query Match 83.9%; Score 1264.5; DB 11; Length 295;
Best Local Similarity 83.9%; Pred. No. 1.4e-98;
Matches 245; Conservative 26; Mismatches 20; Indels 1; Gaps 1;

QY 1 MNTPEPKFGVVVGVGRAGSVRMRLRNPHPSAFNLNLIGFVSRRRLGSDGVQOQISLE 60
|||||
DB 1 MNTPEPKFGVVVGVGRAGSVRIRLDKDPH-SSAFNLNLIGYVSRRELGLDNRQISLE 59

QY 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120
|||||
DB 61 DALSSQVEVAYICSESSSHEDYIRQFLQAGKHVLYEYPMALSFAAQELWELAAQGRV 119

QY 121 LHEHVLLMEEFALFKKVVGVKDLLKGLSLFTAGPLEEERFGPPAFSGISRLTWLSLF 180
|||||
DB 120 LHEHVLLMEEFALFKKVVGVKDLLKGLSLFTAGPLEEERFGPPAFSGISRLTWLSLF 179

QY 181 GELSLVSATLEERKEDQYMKMTVCLETEKKSPLSWIEKGPGLKRNRYLSFHFKSGSLEN 240
|||||
DB 180 GELSLVSATMENKEDQYMKMTVQLETQNKSPISLWIEKGPGLKRNHRIHFKSGSLEE 239

QY 241 VPNVGNKNIFLKDQNTFVQKLLGFSEKELAAEKKRILHCLGLAEIQRKCCSRK 292
|||||
DB 240 VPNVGNKNIFLKDQNTFVQKLLGFSEKELAAEKKRILHCLGLAEIQRKCCSRK 291

RESULT 4
Q9DD21 ID Q9DD21 PRELIMINARY; PRT; 303 AA.
AC Q9DD21;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 0610006A11RIK PROTEIN.
GN 0610006A11RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
```

```
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002231; BAB21950.1; -.
DR MGI; 1915580; 0610006A11Rik.
DR InterPro; IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
SQ SEQUENCE 303 AA; 34491 MW; 52D8A3B02E9556B CRC64;

Query Match 83.9%; Score 1264.5; DB 11; Length 303;
Best Local Similarity 83.9%; Pred. No. 1.4e-98;
Matches 245; Conservative 26; Mismatches 20; Indels 1; Gaps 1;

QY 1 MNTPEPKFVVGVVGRAGSVNRDLRNPSPSAFLNLIGFVSRRELGSIDGVQOISLE 60
Db 1 MSTEPKRFKVVGVVGRAGSVNRDLRNPSPSAFLNLIGFVSRRELGSIDGVQOISLE 59
QY 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120
Db 60 DALRSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 119
QY 121 LHEEHVELLEEFALKKKEVGVKDLKGLSLFTAGPLEERFPGFPAFSGISRLTWLSLF 180
Db 120 LHEEHVELLEEFALKKKEVGVKDLKGLSLFTAGPLEERFPGFPAFSGISRLTWLSLF 179
QY 181 GELSLSVATLEERKEDQYMKMTVCLETEKKSPLSWIEKEGPGFKRNRYLSFHFKSGSLEN 240
Db 180 GELSLSVATLEERKEDQYMKMTVCLETEKKSPLSWIEKEGPGFKRNRYLSFHFKSGSLEN 239
QY 241 VPNGVKNKFLDKDNIFVQKLGQFSEKELAEKRLILHCLGLAEETQKVC 292
Db 240 VPNGVKNKFLDKDNIFVQKLGQFSEKELAEKRLILHCLGLAEETQKVC 291

RESULT 5
Q95019 PRELIMINARY; PRT; 92 AA.
AC Q95019
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BILIVERDIN-IX ALPHA REDUCTASE (FRAGMENT).
GN BLVRA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cordes M., Wohldman P., Morris M.;
RT "The sequence of Homo sapiens PAC clone RP5-974M14.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004939; AAD05025.1; -.
FT NON_TER 1
FT NON_TER 92
SQ SEQUENCE 92 AA; 10552 MW; 390331108449BD93 CRC64;

Query Match 30.7%; Score 463; DB 4; Length 92;
Best Local Similarity 100.0%; Pred. No. 9.7e-32;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 119 KVLHEEHVELLEEFALKKKEVGVKDLKGLSLFTAGPLEERFPGFPAFSGISRLTWLS 178
Db 1 KVLHEEHVELLEEFALKKKEVGVKDLKGLSLFTAGPLEERFPGFPAFSGISRLTWLS 60
QY 179 LFGSLSVATLEERKEDQYMKMTVCLETEKK 210
Db 61 LFGSLSVATLEERKEDQYMKMTVCLETEKK 92

RESULT 6
P72782 PRELIMINARY; PRT; 328 AA.
AC P72782
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BILIVERDIN REDUCTASE.
GN SLR1784.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Saito M., Saito K., Okumura S.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90900; BAA16797.1; -.
DR InterPro; IPR000683; GFO_IDH_MOCA.
DR InterPro; IPR004104; GFO_IDH_MOCA_C.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
DR Pfam; PF02894; GFO_IDH_MOCA_C; 1.
KW Complete proteome.
SQ SEQUENCE 328 AA; 36640 MW; 51474ABD26B2C7EF CRC64;

Query Match 14.4%; Score 217; DB 16; Length 328;
Best Local Similarity 24.9%; Pred. No. 3.2e-10;
Matches 79; Conservative 57; Mismatches 139; Indels 42; Gaps 11;

QY 11 VVVVGVGRAGSVNRDLRNPSPSAFLNLIGFVSRRELGSID----GVQ-QISLEDAIS 64
Db 13 VGVVGTGYAAQRAAEVFRGDRS----QLVSFWGNSEANTAKFADTFCVRPQSQWQALIN 68
QY 65 SQEVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGVLEE 124
Db 69 DPEIDLVLIIATINQHGAIAPAAALQAGKHVLYEYPMTLSLAAQELWELAEQKGVLEE 128
QY 125 HVLEELMEEFALKKKEV--VGKDL-LKGSLLFTAGPLEE-----ERFGFPAFSGISRLTW 175
Db 129 HIELLGVGHQAIRNLGKIGEVFYARYSTIMGQNPAPQRTYTHHQFGFPLVAALSISR 188
QY 176 LVSLFGLSLSVATLE--ERKEDQYMK---MTVCLETEKKSPLSWIEKEGPGFKRN--RYL 229
Db 189 FTDLFGTVQVQVDAOCRFWDQNPVFRACLATAYLQFNNGLKAIEVYKGEVFNQRIIF 248
QY 230 SFHFKSGSL-----ENVPNVGVNKNIFPKDONIFVOKLGLQGFSEKELAEK 275
Db 249 TLHGDRGTLLFVGETGRLLIQGTETITVGSRRGLFRQDTEAVLDYLT---TGKPLVVDL 305
QY 276 KRILHCLGLAEETQKVC 292
Db 306 EASLYALEVADLCAQAC 322

RESULT 7
Q97UHO
```

ID	Q97UH0	PRELIMINARY;	PRT;	371 AA.
AC	Q97UH0;			
AC	01-OCT-2001 (TrEMBLrel. 18, Created)			
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	DEHYDROGENASE, PUTATIVE.			
GN	SSO3049.			
OS	Sulfolobus solfataricus.			
OC	Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.			
OC	NCBI_TaxID=2287;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 35092 / DSM 1617 / P2;			
RX	MEDLINE=21332296; PubMed=11427726;			
RA	She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,			
RA	Wayde M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,			
RA	De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,			
RA	Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,			
RA	Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,			
RA	Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,			
RA	Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;			
RT	"The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).			
DR	EMBL; AE006896; AAK43149.1; -.			
DR	InterPro; IPR000683; GFO_IDH_MoCA.			
DR	InterPro; IPR004104; GFO_IDH_MoCA_C.			
DR	Pfam; PF01408; GFO_IDH_MoCA; 1.			
DR	Pfam; PF02894; GFO_IDH_MoCA_C; 1.			
KW	Complete proteome.			
SK	SEQUENCE 371 AA; 41760 MW; DAAEL6456D53306B CRC64;			
Query Match 10.4%; Score 157.5; DB 17; Length 371;				
Best Local Similarity 19.9%; Pred. No. 3.9e-05;				
Matches 64; Conservative 58; Mismatches 118; Indels 81; Gaps 10;				
Qy	7	RKGVVVVGVGRAGSRMRDLRNPHPSAFUNLIGTVSR-----RELGSIDGV-QQISLE	60	
Db	18	KRIGVAVVGLSGIKTHVKALKDLKETEFVKLVAVVDQIKAIAEKIGSEYGTPTYTTID	77	
Qy	61	DALSSQEVAVYICSESSSHEDIYRQFLNACKGHVLYEYPMTLSLAAAGELWELAEQK---	117	
Db	78	EVLKRSVDVTSIATPSVLHAPQAILAEYKGHVIVEKPMATTLAGAREMVSRAERNEVK	137	
Qy	118	-GKVLHEEH-----VELLMEEF-----AFLKKEVVGKDLLKGLLF	152	
Db	138	LGVIFQERYAPDIRRLKNDILKELGRIVLISELKWYRDMKGYIKRDEIARS-----	189	
Qy	153	TAGPLEERGFPAFGISRLTWLVSFLGELSVLSAYLEERKEOYKMKTVCLTEKKSP	212	
Db	190	WRGMWTEGGGVMTNQGHITDLMWLNGEVEEYSGFVDNLTHDG-----IE-----	236	
Qy	213	LSWTEEGPGKLKRNRYLSHFHPKSGLENPVNV-----GVNKNFIKLDON	256	
Db	237	---VEDTAVALMR-----YKNGALGTISOTVSMKPTTYQYKIRVNGSNGFVEITDGS	286	
Qy	257	IFVQKLLQFSEKELAAEKKR	277	
Db	287	LSTVAIEGKIEESKSSVEYKK	307	
RESULT 8				
P94437				
ID	P94437	PRELIMINARY;	PRT;	393 AA.
AC	P94437;			
DT	01-MAY-1997 (TrEMBLrel. 03, Created)			
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	YFII.			
GN	YFII.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			

RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=AC327;
FX	MEDLINE=971128783; PubMed=8973323;
RA	Yamamoto H., Uchiyama S., Sekiguchi J.;
RT	"The Bacillus subtilis chromosome region near 78 degrees contains the
RT	genes encoding a new two-component system, three ABC transporters and
RT	a lipase.";
RL	Gene 181:147-151(1996).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=168;
RX	MEDLINE=38044033; PubMed=9384377;
RA	Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchart S.,
RA	Boriss R., Bourster L., Brans A., Braun M., Brignell S.C., Bron S.,
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA	Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Fouquier D.,
RA	Fritz C., Fujita M., Fujita Y., Futaba S., Galizzi A., Galleron N.,
RA	Ghim S.Y., Glaser P., Goiffeau A., Golightly E.J., Grandi G.,
RA	Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA	Hilbert H., Holtsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
RA	Joris B., Karanata D., Kasahara Y., Klaerner-Blandhard M., Klein C.,
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA	Kurita K., Lapidos A., Lardinou S., Lauber J., Lazarevic V.,
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA	Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA	Parron V., Poh T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA	Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA	Sekiguchi J., Sekowska A., Seror S.J., Serrif P., Shin B.S., Solido B.
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA	Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognohi A.,
RA	Tosato V., Uchiyama S., Vandenberg M., Vanier F., Vassarotti A.,
RA	Vitari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA	Winters P., Wipat A., Yamamoto H., Yanane K., Yasumoto K., Yata K.,
RA	Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT	The complete genome sequence of the gram-positive bacterium Bacillus
RT	subtilis."
Nature	390:249-256(1997).
RL	Nature 390:249-256(1997).
[3]	
NP	SEQUENCE FROM N.A.
RC	STRAIN=168;
RA	Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL	Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR	EMBL: D78508; BAAL1399.1 ; -
DR	ENBL: Z99108; CABI2657.1 ; -
DR	InterPro: IPR000683; GFO_IDH_Moca_C.
DR	InterPro: IPR004104; GFO_IDH_Moca_C.
DR	Pfam: PF01408; GFO_IDH_Moca_1.
DR	Pfam: PF02894; GFO_IDH_Moca_C_1.
KW	Complete proteome.
SQ	SEQUENCE 393 AA; 43584 MW; 3EEB326E1A047602 CRC64;
	Query Match 9.5%; Score 143; DB 16; Length 393;
	Best Local Similarity 24.4%; Pred.No. 0.0007;
	Matches 66; Conservative 52; Mismatches 97; Indels 56; Gaps
Qy	13 VVGVRAGSV----RMRDLRNHPSSAF-LNLIGFVSRRRELGSIDGV----- 54 :: : ::
Db	17 MVGGGRLSQGKYKRGIRGALKRD---NTAQLTAGATFDIAERGDKGFCNVGLVDARCYPNW 73 ::
Qy	55 QQLSIEDALSQQEVAYICSESSHEDIYRFQFNAGKHVLVEYPMTLSLAQAELWEILA 114 : -:- ::
Db	74 QTFAFEAEAKQDGIETVYSIAPNGTHYEICKALEAGVHVHCPEKLPIFFTSAGEQEIKALA 133 : -:- ::
Qy	115 EQKKVKLVHEH----VELLMEEFAFLUKKEVVGVKDILKLGSLITTAG---PLEBERFG-- 164 : : : :- : : : : : : : :
Db	134 EKKGKIIVGYTFSGNMLOMLOMEAMIEOGMI-DIRWDLDVTHTGCARDDEGEKISAAOK 192 : -:- ::

[illegible]

```

RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambutt R., Weltzenegger T., Pohl T.M., Terryn N.,
RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA Entong K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milloni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Reichman S., Ansoorge W., Cooke R., Berger C.,
RA Delseny M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chalwatzis N.;
RA "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana.";
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97343; CAB10518.1; -.
DR EMBL; AL161546; CAB78740.1; -.
DR InterPro; IPR000683; GFO_IDH_Moca.
DR InterPro; IPR004104; GFO_IDH_Moca.C.
DR Pfam; PF01408; GFO_IDH_Moca; 1.
DR Pfam; PF02894; GFO_IDH_Moca.C; 1.
DR PF02894; GFO_IDH_Moca.C; 1.
SQ SEQUENCE 364 AA; 40260 MW; B56312BD6CF01B95 CRC64;

Query Match 9.1%; Score 136.5; DB 10; Length 364;
Best Local Similarity 23.1%; Pred. No. 0.0022;
Matches 61; Conservative 49; Mismatches 89; Indels 65; Gaps 13;

QY 8 KFGVVVVG-VGRAGSVRMRDLR-----NHPSSAFLNL-----IGFVSRRELGSI 51
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 KYGIVGIMMGREHLNLHLRDOGLAVVCIADPHPPSOLLALIAEQSFGW----ELKVF 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 52 DGVOQISLEDALSSOEVEVAYICSESSSHEDYIROFLNAGK--HVLVEYPMTLSLAARQE 109
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 SG-----HEELKSELCDVIVVSSPNMTHQILMDIINYKPHHVLVEKPLCTTVADCKQ 121
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 110 LWELAEQKGVKLHHEHVEL-LMEEFALFKKEVGKDLKLSLLFTAGPLEERFGFPAPS 168
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 VLEAAKRSDMVQVGLERYVWPVAKLIEQVKGRDF--GNVKVA--IREHPPFLVKV 177
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 169 GISRL-TW-----LVSLFGEISLVSATLEERKEDQYMKMTVCLETKSKPSLSWI 216
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 GLNKMLGLTWLVEKCHFPDLRLFAGAN-----PVCVMAAGGMDVNHK 220
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 217 EE---KPGPLKRNRYLSFHKSG 236
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 DEVYGGKVPDIIDNAYVIEFDNG 244
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
Q9WZ54 PRELIMINARY; PRT; 360 AA.
ID Q9WZ54 AC
AC Q9WZ54;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LIPOPOLYSACCHARIDE BIOSYNTHESIS PROTEIN BPLA.
GN TM0585.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OC NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Matek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;

```

RT	"Evidence for lateral gene transfer between Archaea and Bacteria from		
RL	genome sequence of <i>Thermotoga maritima</i> ."		
DR	EMBL; AE001733; AAD35670.1; -		
DR	TIGR; TM0585; -		
DR	InterPro; IPR000683; GFO_IDH_MoCA.		
DR	InterPro; IPR004104; GFO_IDH_MoCA_C.		
DR	InterPro; IPR003489; Ribosomal_S30.		
DR	Pfam; PF01408; GFO_IDH_MoCA; 1.		
DR	Pfam; PF02894; GFO_IDH_MoCA_C; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 360 AA; 41023 MW; 3B17BCD4A94A6EC6 CRC64;		
	Query Match	8.8%; Score 132; DB 16; Length 360;	
	Best Local Similarity	23.0%; Pred. No. 0.0053;	
	Matches 65; Conservative 37; Mismatches 104; Indels 76; Gaps		
Qy	8 KFGVVVGVGRAGS-----VNRDLRNPHPSAFUNLIGFVSRRELGSID--GVQO--	56	
Db	2 KLRIALVCGGRIGQKHKVPAIIETQDL---FETVAVCDLVEERANRAABHFKEKSLRRPE	58	
Qy	57 --ISLEDALSSQEEVAVYICSESSSHEDYIQFLNAGKQHLVVEYPTWLSIAAAQELWELA	114	
Db	59 TWTNYRELLKREDVDVTSIAFESGKHVQITMEALEAGKQHLVVEKPKALSTKHNMVVELS	118	
Qy	115 BQKQVLHEEHVELLEEFALKEKVVGVKDLLKGSLLFTAGPLEEERFGPPAFSGISRLT	174	
Db	119 KQKN-----LKLGVFF-----QNRPNPQVEVRKKLD	145	
Qy	175 WLVSFLGELSIVSATLEERKEDQYMKMTVCLETETKKSPLSW--IEKGPGLKRNRYLSPH	232	
Db	146 --SGAFGKIFYASVAVWRNRNEDYKQA-----SWRGTMWDGGVLMNQ--STH	190	
Qy	233 -----FKSGSLENVPNGVKNK---IFLKDQNIFFVKLLG	264	
Db	191 AIDLQWFLGGEIEIYGHIANTHNPYIEADEGFAIVKPKG	232	
RESULT 13			
Q9VQB3	PRELIMINARY; PRT; 335 AA.		
AC	Q9VQB3		
DT	01-MAY-2000 (TremBLrel. 13, Created)		
DT	01-MAY-2000 (TremBLrel. 13, Last sequence update)		
DE	01-DEC-2001 (TremBLrel. 19, Last annotation update)		
DE	CG3597 PROTEIN.		
GN	CG3597.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,		
RA	George R.A., Lewis S.E., Richards S.D., Vandell M.D., Zhang Q., Chen L.X.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Blazek R.G., Champe M., Pfeiffer B.D.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,		
RA	Arbail J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		

```
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003583; AF51263.1;
DR FlyBase: FBgn0031417; CG3597.
DR InterPro: IPR000683; GFO_IDH_MoCA_C.
DR InterPro: IPR004104; GFO_IDH_MoCA_C.
DR Pfam: PF01408; GFO_IDH_MoCA_C.
DR Pfam: PF02894; GFO_IDH_MoCA_C.
DR SEQUENCE 335 AA; 37812 MW; ELAEB59E5F270612 CRC64;

Query Match 8.7%; Score 130.5; DB 5; Length 335;
Best Local Similarity 23.2%; Pred. No. 0.0064;
Matches 71; Conservative 43; Mismatches 96; Indels 96; Gaps 15;

Qy 1 MNTPEPKFVVVG-----VGRAGSVRMEDLNRPSSAFNLNIG-----FVSRREL- 48
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 1 MQSQANLWGAAGRITQDVTALGTVE---KSRHVVAADVQDQRAQQFAQRNP 56
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 49 GSIDQVOQISLEDALSSQEVAVICSESSHEDYIROFLNAGKHLVPEYPTLSLAAQ 108
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 57 RHYDGFDAID-----REVVVVGTLPNPHYAVVHLMARGKNVLCETPMCLSVQAK 111
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 109 ELWELAEQKQVLHBEHVLLMEEF---AFLKKEVVGK-----D 144
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 112 ELYTLAEQRGVFLMEGMWSRFFPSYDRLLKNDVIGEVTVQVKVGHGRLAHMERVCNR 171
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 145 LKGSLLFTAG--PLEERFGFPFAGSIS-----GVSVPKILPSGTQLNKERVQIDFMDYGDGRR 227
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 172 SLGGSILMDIGIYALQIQGFV---GVSPVKILPSGTQLNKERVQIDFMDYGDGRR 227
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 173 LTWLVSIFGEL---SLVSATLEERKEDQYMKMTVCLETEKKS---PLSWIEKGPGLKRN 227
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 228 LVALVTGLNLENDAVITGTGKELSNW---CCTQISRSNAPPESW-----PLPRAK 278
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 228 YLSFHF 233
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 279 F-DFHY 283

RESULT 14
O42896
ID O42896 PRELIMINARY; PRT; 368 AA.
AC O42896;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE C115.03 IN CHROMOSOME II (EC 1.-.-.-).
GN SPBC115.03 OR SPBC24E9.18C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Devlin K., Churcher C.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE YHx FAMILY OF OXIDOREDUCTASES.
DR EMBL: AL022105; CAAL17915.1;
DR InterPro: IPR000683; GFO_IDH_MoCA_C.
DR InterPro: IPR004104; GFO_IDH_MoCA_C.
DR Pfam: PF01408; GFO_IDH_MoCA_C.
DR Pfam: PF02894; GFO_IDH_MoCA_C.
DR Pfam: PF02894; GFO_IDH_MoCA_C.
KW Hypothetical protein; Oxidoreductase.
FT DOMAIN 359 365 POLY-SER.
SQ SEQUENCE 368 AA; 41271 MW; D2DC7A6AA2C3F377 CRC64;

Query Match 8.7%; Score 130.5; DB 3; Length 368;
Best Local Similarity 24.1%; Pred. No. 0.0073;
Matches 45; Conservative 30; Mismatches 61; Indels 51; Gaps 4;

Qy 34 SAFNLNIGFVSRRELGSIDGVQOQISLEDALSSQEVAVICSESSHEDYIROFLNAGKH 93
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 49 SAYPNILVYT-----NLDELLADVIELVVVSLPPNVHSEIVKKALNACKH 94
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 94 VLVEYPTLSLAAQELWELAEQKQVLHBEHVLLMEEFAPFLKKEVVGKDLLKGSLLPT 153
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 95 VVCEKPEPTTYEAKELYELAESKSLI-----AIYQNRWRWDGDLTAKKIE 142
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 154 AGPL-----EERFGFPFAGSISR--LTWLVSIFGELSLSVA 188
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 143 SGRIGQVVVFESHFDYRLGRKSGWKDEPRGNGMVGSHLIDQAVSLFGTPYSVTA 202
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 189 TLEERKE 195
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 203 KLEAQRQ 209

RESULT 15
O9CF67
ID O9CF67 PRELIMINARY; PRT; 324 AA.
AC O9CF67;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE OXIDOREDUCTASE.
GN YPJF.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403.
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006386; AAK05652.1;
DR InterPro: IPR000683; GFO_IDH_MoCA_C.
DR Pfam: PF01408; GFO_IDH_MoCA_C.
DR Complete proteome.
SQ SEQUENCE 324 AA; 36603 MW; DBF47D66684EB259C CRC64;

Query Match 8.6%; Score 129.5; DB 16; Length 324;
Best Local Similarity 21.5%; Pred. No. 0.0074;
Matches 64; Conservative 45; Mismatches 88; Indels 101; Gaps 13;

Qy 8 KFGVV-----VVGVRAGSVRMEDLNRPSSAFNLNIGFVSRRELSDGVQOIS 58
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Db 3 KIGINGAAKIVPRFVAGVKESQA-----EVTGIAARNK-----EKAQKAA 43
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
Qy 59 LE-----DALSSQEVAVICSESSHEDYIROFLNAGKHLVPEYPTLSLAA 107
| : : : : : | : : : : : | : : : : : | : : : : : | : : :
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:41:32 ; Search time 15.2283 Seconds
(without alignments)
752.611 Million cell updates/sec

Title: US-09-606-129a-3

Perfect score: 1508

Sequence: 1 MNTEPRKGVVVGVVGRAG.....RILHCLGLAEIQKYCCSRK 296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1490	98.8	296	1 BIEA_HUMAN	P53004 homo sapien
2	1247.5	82.7	295	1 BIEA_RAT	P46844 rattus norv
3	151.5	10.0	345	1 YHHX_ECOLI	P46853 escherichia
4	130.5	8.7	371	1 YH16_SYNY3	P74041 synechocyst
5	122.5	8.1	358	1 YVAA_BACSU	O32223 bacillus su
6	119.5	7.9	328	1 YULF_BACSU	O20265 bacillus su
7	108	7.2	341	1 YRBE_BACSU	O05389 bacillus su
8	105.5	7.0	349	1 YM94_YEAST	O04869 saccharomyc
9	104	6.9	334	1 YMO1_RHIME	P49305 rhizobium m
10	103.5	6.9	397	1 CBAC_COMTE	O44258 comamonas t
11	101.5	6.7	439	1 GFO_ZYMMO	O07982 zymomonas m
12	100	6.6	307	1 MVIM_ECOLI	P75931 escherichia
13	100	6.6	346	1 YDGJ_ECOLI	P77376 escherichia
14	100	6.6	410	1 PHT4_PSEPU	O05184 pseudomonas
15	100	6.6	533	1 ANFA_AZOVI	P12626 azotobacter
16	99.5	6.6	1076	1 RPOB_ASTLO	P27059 astasia lon
17	99	6.6	342	1 YISS_BACSU	P40332 bacillus su
18	98.5	6.5	1316	1 RPOC_MYCTU	P47769 mycobacteri
19	98	6.5	579	1 G160_HUMAN	O08378 homo sapien
20	98	6.5	2418	1 SPCA_HUMAN	P02549 homo sapien
21	97	6.4	988	1 TNP6_ENTFC	O06238 enterococcu
22	95.5	6.3	2230	1 GOG4_HUMAN	O13439 homo sapien
23	95	6.3	328	1 YGJR_ECOLI	P42599 escherichia
24	94.5	6.3	1548	1 SMCV_MOUSE	O62240 mus musculu
25	94	6.2	316	1 YH58_AQUAE	O67638 aquifex aeo
26	94	6.2	931	1 DING_BACSU	P54394 bacillus su
27	93	6.2	307	1 MVIM_SALTY	P37168 salmonella
28	92.5	6.1	545	1 FLIF_BUCAI	P57175 buchnera ap
29	92.5	6.1	2051	1 FAS1_YEAST	P07149 s fatty aci
30	92	6.1	330	1 M12D_RHIME	O68965 rhizobium m
31	91.5	6.1	1959	1 MYH9_CHICK	P14105 gallus galli
32	91	6.0	695	1 VATI_METJA	O57675 methanococc
33	90.5	6.0	724	1 HMWR_HUMAN	O75330 homo sapien

RESULT 1	BIEA_HUMAN	STANDARD;	PRT;	296 AA.
ID	BIEA_HUMAN			
AC	P53004;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Biliverdin reductase A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-reductase).			
DE	BLVRA OR BLVR OR BVR.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=96202961; PubMed=8631357;			
RA	Maines M.D., Polevoda B.V., Huang T.-J., McCoubrey W.K. Jr.;			
RT	"Human biliverdin IXalpha reductase is a zinc-metalloprotein.			
RT	Characterization of purified and Escherichia coli expressed			
RT	enzymes";			
RL	Eur. J. Biochem. 235:372-381(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Komuro A., Tobe T., Nakano Y., Yamaguchi T., Tomita M.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-117 FROM N.A.			
RA	Cordes M., Wollam C., Carter T.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE OF 3-36; 48-74 AND 228-248.			
RC	TISSUE=Liver;			
RX	MEDLINE=93143333; PubMed=8424666;			
RA	Maines M.D., Trakshel G.M.;			
RT	"Purification and characterization of human biliverdin reductase.";			
RL	Arch. Biochem. Biophys. 300:320-326(1993).			
RN	[5]			
RP	SEQUENCE OF 3-22.			
RC	TISSUE=Liver;			
RX	MEDLINE=95014177; PubMed=7929092;			
RA	Yamaguchi T., Komoda Y., Nakajima H.;			
RT	"Biliverdin-IX alpha reductase and biliverdin-IX beta reductase from			
RL	human liver. Purification and characterization.";			
RL	J. Biol. Chem. 269:24343-24348(1994).			
CC	- - FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN.			
CC	- - CATALYTIC ACTIVITY: Bilirubin + NAD(P)(+) = biliverdin + NAD(P)H.			
CC	- - COFACTOR: BINDS ONE ZINC ION. HAS DUAL PH/COFACTOR (NADH, NADPH)			
CC	- - SPECIFICITY. USES NADH AT THE ACIDIC PH RANGE (6-6.7) AND NADPH AT			
CC	THE ALKALINE RANGE (8.5-8.7).			
CC	- - PATHWAY: FINAL STEP IN HEME METABOLISM.			
CC	- - SUBUNIT: MONOMER.			
CC	- - SUBCELLULAR LOCATION: Cytoplasmic.			
CC	- - TISSUE SPECIFICITY: LIVER.			
CC	- - SIMILARITY: TO E. COLI YHHX.			

P27725 serratia ma
O9m714 nicotiana t
O51555 borrelia bu
P42499 glycine max
P55937 mus musculu
P40457 saccharomyc
P39353 escherichia
O62209 mus musculu
O01550 xenopus lae
P26370 saccharomyc
P13925 streptococc
O58451 pyrococcus

ALIGNMENTS

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; X93086; CAA63635.1; -;
 DR EMBL; U34877; AAC35588.1; -;
 DR EMBL; AC005189; AAC25526.1; -;
 DR MIM; 109750; -;
 DR InterPro; IPR000683; GFO_IDH_MocA.
 DR Pfam; PF01408; GFO_IDH_MocA; 1.
 KW Oxidoreductase; NAD; NADP; Zinc.
 FT PROPEP 1 2
 FT CHAIN 3 296 BILIVERDIN REDUCTASE A.
 FT DOMAIN 11 16 POLY-VAL.
 FT METAL 280 280 ZINC (POTENTIAL).
 FT METAL 281 281 ZINC (POTENTIAL).
 FT METAL 292 292 ZINC (POTENTIAL).
 FT METAL 293 293 ZINC (POTENTIAL).
 FT CONFLICT 3 3 A -> T (IN REF. 2).
 FT CONFLICT 154 155 SD -> AG (IN REF. 2).
 FT CONFLICT 160 160 D -> E (IN REF. 2).
 SQ SEQUENCE 296 AA; 33488 MW; 0DFD3B386F4DFC0A CRC64;

 Query Match 98.8%; Score 1490; DB 1; Length 296;
 Best Local Similarity 98.8%; Pred. No. 1.9e-108;
 Matches 292; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNTPEPKFGVVGVRAGSVRMRLRNPPSSAFNLIGFVSRRELGSIDGVQOISLE 60
 DB 1 MNAEPKFKGVVGVRAGSVRMRLRNPPSSAFNLIGFVSRRELGSIDGVQOISLE 60
 QY 61 DALSSQSEVAYICSSSSSHEDYIROFLNAGKHVLYVEYPMTLSLAAQELWELAEQKGV 120
 DB 61 DALSSQSEVAYICSSSSSHEDYIROFLNAGKHVLYVEYPMTLSLAAQELWELAEQKGV 120
 QY 121 LHEEHVELLMEEFAFLKKEVGVGKDLKGLLFTAGPLEEERFGFPAPSGISRLTWLVSLF 180
 DB 121 LHEEHVELLMEEFAFLKKEVGVGKDLKGLLFTAGPLEEERFGFPAPSGISRLTWLVSLF 180
 QY 181 GELSLVSATLEERKEDQYKMTVCLETKKSPLSWIEKPGGLKRNRYLSFHFKSGSLEN 240
 DB 181 GELSLVSATLEERKEDQYKMTVCLETKKSPLSWIEKPGGLKRNRYLSFHFKSGSLEN 240
 QY 241 VPVGVNKNIFLKDQNFVQKLLGFQSEKELAAEKKRILHCLGLAEIEIQKYCCSRK 296
 DB 241 VPVGVNKNIFLKDQNFVQKLLGFQSEKELAAEKKRILHCLGLAEIEIQKYCCSRK 296

RESULT 2
 BIEA_RAT
 ID BIEA_RAT STANDARD; PRT; 295 AA.
 AC P46844;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 13-JUL-1999 (Rel. 38, Last annotation update)
 DE Biliiverdin reductase A precursor (EC 1.3.1.24) (Biliiverdin-IX alpha-
 DE reductase).
 DE BLVRA OR BLVR.
 GN Rattus norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Kidney;
 RX MEDLINE=92156147; PubMed=1371282;
 RA Fakhrat H., Maines M.D.;
 RT "Expression and characterization of a cDNA for rat kidney biliiverdin

RT reductase. Evidence suggesting the liver and kidney enzymes are the
 RT same transcript product.";
 RL J. Biol. Chem. 267:4023-4029(1992).
 RN [2]
 RP MUTAGENESIS:
 RX MEDLINE=94291657; PubMed=8020496;
 RA McCoubrey W.K. Jr., Maines M.D.;
 RT "Site-directed mutagenesis of cysteine residues in biliiverdin
 RT reductase. Roles in substrate and cofactor binding.";
 RL Eur. J. Biochem. 222:597-603(1994).
 CC -|- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN; DISPLAYS TWO DISTINCT
 CC PH OPTIMA USING A DIFFERENT COFACTOR AT EACH PH: NADH AT THE LOWER
 CC PH 6.7-6.9 RANGE AND NADPH AT PH 8.5-8.7. NADPH, HOWEVER, IS THE
 CC PROBABLE COFACTOR IN BIOLOGICAL SYSTEMS.
 CC -|- CATALYTIC ACTIVITY: Biliirubin + NAD(P)(+) = biliiverdin + NAD(P)H.
 CC -|- COFACTOR: BINDS ONE ZINC ION.
 CC -|- PATHWAY: FINAL STEP IN HEME METABOLISM.
 CC -|- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -|- SIMILARITY: TO E.COLI YHHX.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; M81681; AAA40830.1; -;
 DR InterPro; IPR000683; GFO_IDH_MocA.
 DR Pfam; PF01408; GFO_IDH_MocA; 1.
 KW Oxidoreductase; NAD; NADP; Zinc.
 FT PROPEP 1 2
 FT CHAIN 3 295 BILIVERDIN REDUCTASE A.
 FT DOMAIN 11 16 POLY-VAL.
 FT METAL 279 279 ZINC (POTENTIAL).
 FT METAL 280 280 ZINC (POTENTIAL).
 FT METAL 291 291 ZINC (POTENTIAL).
 FT METAL 292 292 ZINC (POTENTIAL).
 FT MUTAGEN 73 73 C->A: LOSS OF ACTIVITY.
 FT MUTAGEN 280 280 C->A: REDUCED ACTIVITY.
 FT MUTAGEN 291 291 C->A: REDUCED ACTIVITY.
 SQ SEQUENCE 295 AA; 33565 MW; 219C8EA96C150588 CRC64;

 Query Match 82.7%; Score 1247.5; DB 1; Length 295;
 Best Local Similarity 82.8%; Pred. No. 1.2e-89;
 Matches 245; Conservative 26; Mismatches 24; Indels 1; Gaps 1;
 QY 1 MNTPEPKFGVVGVRAGSVRMRLRNPPSSAFNLIGFVSRRELGSIDGVQOISLE 60
 DB 1 MDAEPKFKGVVGVRAGSVRLRLKDPK-SAAFNLIGFVSRRELGSIDGVQOISLE 59
 QY 61 DALSSQSEVAYICSSSSSHEDYIROFLNAGKHVLYVEYPMTLSLAAQELWELAEQKGV 120
 DB 60 DALRSQEIDVAYICSSSSSHEDYIROFLQAGKHVLYVEYPMTLGFAAAQELWELAAQKGV 119
 QY 121 LHEEHVELLMEEFAFLKKEVGVGKDLKGLLFTAGPLEEERFGFPAPSGISRLTWLVSLF 180
 DB 120 LHEEHVELLMEEFEURREYVGLKGLSLRFTASPLEEERFGFPAPSGISRLTWLVSLF 179
 QY 181 GELSLVSATLEERKEDQYKMTVCLETKKSPLSWIEKPGGLKRNRYLSFHFKSGSLEN 240
 DB 180 GELSLVSATLEERKEDQYKMTVCLETKKSPLSWIEKPGGLKRNRYLVNFQFTSSLEE 239
 QY 241 VPVGVNKNIFLKDQNFVQKLLGFQSEKELAAEKKRILHCLGLAEIEIQKYCCSRK 296
 DB 240 VPSGVNKNIFLKDQNFVQKLLDQVSAEDLAAAEKKRIMHCLGLASDIQKLCQHK 295

 RESULT 3
 YHHX_ECOLI
 ID YHHX_ECOLI STANDARD; PRT; 345 AA.

```
AC P46853;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative oxidoreductase yhhX (EC 1.-.-.-).
GN YHHX OR B3440.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaser J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=99420866; PubMed=10493123;
RA Pounoukakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195 (1999).
CC -!- SIMILARITY: BELONGS TO THE YHHX FAMILY OF OXIDOREDUCTASES.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U18997; AAA58238.1; -.
DR EMBL; AE000420; AAC76465.1; -.
DR EcoGene; EG12944; yhhX.
DR InterPro; IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA.1.
DR Pfam; PF02894; GFO_IDH_MOCA.C; 1.
DR Oxidoreductase; Complete proteome.
SQ SEQUENCE 345 AA; 38765 MW; 31D73A62511C7F68 CRC64;

Query Match 10.0%; Score 151.5; DB 1; Length 345;
Best Local Similarity 31.3%; Pred. No. 0.00012;
Matches 42; Conservative 27; Mismatches 52; Indels 13; Gaps 4;

QY 59 LEDALSSQVEVAVICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAQKG 118
Db LDEVLNDPDKLVVYKTHADSHFEYAKRALEAGKNVLVEKPFPTTQAQKELFALAKSG 116
QY 119 KVLHEEHVELLMERFAFLKEWGDLLKGLSLFTAG-----PLEERFGFP---AF-- 167
Db LTVTYPQNRFRDSCFLAKKAIESKGL--GEIVEVESHDFYRVPVETKPLPDGAFYG 174
QY 168 SGISRLTLWLSLFG 181
Db LGVHTMQIIISLFG 188

RESULT 4
Y816_SYNY3
ID Y816_SYNY3 STANDARD; PRT; 371 AA.
AC P74041;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical oxidoreductase sll0816 (EC 1.-.-.-).
GN SLL0816.
```

```
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res 3:109-136 (1996).
CC -!- SIMILARITY: BELONGS TO THE GFO_IDH/MOCA FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D90911; BAA18116.1; -.
DR InterPro; IPR000683; GFO_IDH_MOCA.
DR InterPro; IPR004104; GFO_IDH_MOCA.C.
DR Pfam; PF01408; GFO_IDH_MOCA.1.
DR Pfam; PF02894; GFO_IDH_MOCA.C; 1.
DR Hypothetical protein; Oxidoreductase; Complete proteome.
SQ SEQUENCE 371 AA; 41489 MW; BDEFE020D4B84BF6 CRC64;

Query Match 8.7%; Score 130.5; DB 1; Length 371;
Best Local Similarity 26.0%; Pred. No. 0.0056;
Matches 47; Conservative 33; Mismatches 56; Indels 45; Gaps 7;

QY 11 VVVGVGRAGSRMRDLNRNPPSSAFLNLIGFVSRRRLSGIDGVQOI-----SLEDA 62
Db VAVIGTGFQGVAVHIPALQYHQQTQIA-----IYHRLAKAQEYAKSNDLAYSYNNLEEL 63
QY 63 LSSQVEVAVICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAQKGKVLH 122
Db LANPEQAVTITASPFLHYEMAKQAILAGKHVLEKPMPLRVEETIELYHLARQR----- 118
QY 123 EEHVVELLME-EFAFLK-----KEVVGDLLKGLSLFTAGPLEERFGFPAPSGISRLTWL 176
Db L---:---:---:---:---:---:---:---:---:---:---:---:---:---:---: 119 --EVQVIPDFEFVPAWQVYVALLGQIL-----GOLK-----LIKVDWL 157
QY 177 V 177
Db 158 V 158

RESULT 5
YVAA_BACSU
ID YVAA_BACSU STANDARD; PRT; 358 AA.
AC O52223;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical oxidoreductase in FHUO-OPUD intergenic region.
GN YVAA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OX Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE YHHX FAMILY OF OXIDOREDUCTASES.
```


Best Local Similarity 21.9%; Pred. No. 0.45;
Matches 47; Conservative 38; Mismatches 87; Indels 43; Caps 6;

QY	58	SLEDALSSQBEVAYITCSESSSHEDYTRQFLNACKGHVLYEYPMWTLSLAAQOELWELAEQK	117
DB	60	NLDEILNDPHVDYIDALLPAQFNADIVEKAVKAGPVLIEKPIAINDQAKEIVKIAEST	119
QY	118	GK-----VLHEEHVELLMEEFAFLKKEVVGKDLLKGLSLFTAGPLEEERF-	162
DB	120	PLPVGVAENWLYLPCKIAEQEKGPVAVFTHNS-TGPFVTONKYLTTTWRQKPEHG	178
QY	163	GFPAFSGISRLTDLVSLFGSLVSATLEBKE-----DQYMKMTVCLETEKKSPLSWIEE	218

```

179  GFESDQVRGRLNLSLDSFGSGYSALIRNVRKNSGADLLAFVAVQL
223
QY 219  KGPLKRNRYLSFHFSGSGLENSVPNGVGNKNIFLK 253
      : : : | |
      : : : | |
Db 226  -----KNKEVIGSFTYGSAGF-----ATEKSVFLK 250

RESULT 9
YMOI_RHIME
ID YMOI_RHIME STANDARD; PRT; 334 AA.
AC P49305;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypochemical 36.4 kDa protein in mocC-mocA intergenic region (ORF334).
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN 111

```

```

CC STRAIN=LJ-30;
MEDLINE=95147842; PubMed=7845353;
RX Rossbach S., Kulp D.A., Rossbach U., de Bruijn F.J.;
RT "Molecular and genetic characterization of the rhizopine catabolism
RA (mcaABC) genes of Rhizobium meliloti L5-30.";
RT Mol. Gen. Genet. 245:11-24(1994).
RL -I- SIMILARITY: BELONGS TO THE GFO>IDH/MOCA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/or
CC or send an email to licensese@isb-sib.ch). CC -----
DR EMBL: X78503; CAA55267.1; -.
DR InterPro: IPR000683; GFO_IDH_MOCA.
DR InterPro: IPR004104; GFO_IDH_MOCA_C.
DR Pfam: PF01408; GFO_IDH_MOCA_1.
DR Pfam: PF02894; GFO_IDH_MOCA_C; 1.
KW Hypothetical protein; Oxidoreductase.
SQ SEQUENCE 334 AA; 36408 MW; 1B26045491A5AFFC CRC64;
Query Match 6.9%; Score 104; DB 1; Length 334;
Best Local Similarity 22.2%; Pred. No. 0.56;
Matches 48; Conservative 26; Mismatches 74; Indels 68; Gaps 7;
```

QY	33	SSAF	LMHGVSK	-----	RELSSLDGVQSLZLRUSQZVEVAILGCS	77	
Db	25	SAGLEV	VATASRDLDRARAAT	TRFGISYGSYD	-----BILADPEIEAVYIPLN	76	
QY	78	SSHEDY	TRQFLNAGKHYLV	VEYPMT	LSIAAAGELWELAEOKGVUHE	123	
Db	77	HLHWHA	IRAAEAGKHVLC	EKPLADV	VEELSRLIDCRDTRGRITQEAVM	136	
QY	124	-----	EHVELMBE	FAFLUKKEV	GWKDLKLGSLITFTAG	-----164	
Db	137	DIVASGEI	GEVRAIQGVY	TEVNDLPK	STVNDASIGGGALYDLGVY	PTAAASRMVFAABEP	196

[illegible]

```
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative 4,5,-dihydroxyphthalate dehydrogenase (EC 1.-.-.-)
DE (DHP dehydrogenase).
GN PHT4.
OS Pseudomonas putida.
OG Plasmid PHT.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMH102-2;
RA Nomura Y., Nakagawa M., Ogawa N., Harashima S., Oshima Y.;
RT "Genes in PHT plasmid encoding the initial degradation pathway of
RL phthalate in Pseudomonas putida.";
RL J. Ferment. Biogeng. 74:333-344(1992).
CC -1- FUNCTION: TRANSFORMS 4,5-DIHYDRO-4,5-DIHYDROXYPHTHALATE TO
CC 4,5-DIHYDROXYPHTHALATE.
CC -1- PATHWAY: SECOND STEP OF PHTHALATE DEGRADATION.
CC -1- INDUCTION: INDUCED BY PHTHALATE AND REPRESSED BY GLUCOSE.
CC -1- SIMILARITY: TO ALCALIGENES SP. 1-CARBOXY-3-CHLORO-3,4-
CC DIHYDROXYCYCLOHEXA-1,5-DIENE DEHYDROGENASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13229; BAA02512.1; -
CC InterPro: IPR000683; GFO_IDH_MOCA.
CC InterPro: IPR004104; GFO_IDH_MOCA.C.
CC Pfam; PF01408; GFO_IDH_MOCA; 1.
CC Pfam; PF02894; GFO_IDH_MOCA.C; 1.
CC Plasmid; Oxidoreductase; Aromatic hydrocarbons catabolism.
KW Plasmid; Oxidoreductase; Aromatic hydrocarbons catabolism.
SQ SEQUENCE 410 AA; 45644 MW; 0D1A95B019DCACIA CRC64;

Query Match 6.6%; Score 100; DB 1; Length 410;
Best Local Similarity 29.2%; Pred. No. 1.5;
Matches 33; Conservative 15; Mismatches 47; Indels 18; Gaps 3;

Qy 2 NTEPERKGVVVGVGRAGSVR---MRDLRNPSPSAFLNLIGFSRRRLG-----SI 51
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 HAESTARLGLGVGLGRAFTLMLTFLADRR-----VLGVGACDPREQARQFERDF 56
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 52 DGVOQISLEDALSSQEVAVICSESSHEDYIRQFLNAGKHVLVEYPMTLSL 104
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 DAPAYETIEDLADSNVDALYIASPHQFAEHTRIAAANKRHVLVEXPMALSL 109
: : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
ANFA_AZOVI STANDARD; PRT; 533 AA.
ID ANFA_AZOVI
AC P12626;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Nitrogen fixation protein anfa.
GN ANFA.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89255091; PubMed=2722750;
RA Joergers R.D., Jacobson M.R., Bishop P.E.;
RT "Two nifA-like genes required for expression of alternative
RT nitrogenases by Azotobacter vinelandii.";
RL J. Bacteriol. 171:3258-3267(1989).
CC -1- FUNCTION: ANFA IS ESSENTIAL FOR NITROGEN FIXATION UNDER MO- AND
```

```
CC V-DEFICIENT CONDITIONS. IT IS REQUIRED FOR THE REGULATION OF
CC NITROGENASE 3 TRANSCRIPTION. INTERACTS WITH SIGMA-54.
CC -1- SIMILARITY: 37% IDENTITY TO AZOTOBACTER VINELANDII NIFA AND 61%
CC IDENTITY TO AZOTOBACTER VINELANDII VNEA.
CC -1- SIMILARITY: THE CENTRAL REGION CONTAINS A SIGMA-54 FACTOR
CC INTERACTION ATP-BINDING DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M06751; AAA22146.1; -
CC PIR; A44514; A44514.
CC InterPro: IPR003018; GAF.
CC InterPro: IPR002197; HTH_Fis.
CC InterPro: IPR002078; Sig54_Interact...
CC Pfam; PF01590; GAF; 1.
CC Pfam; PF02954; HTH_8; 1.
CC Pfam; PF00158; Sigma54_activat; 1.
CC SMART; SM00065; GAF; 1.
CC PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
CC PROSITE; PS00675; SIGMA54_INTERACT_2; 1.
CC PROSITE; PS00688; SIGMA54_INTERACT_3; 1.
CC PROSITE; PS00688; SIGMA54_INTERACT_4; 1.
CC PROSITE; PS50045; SIGMA54_INTERACT_4; 1.
KW Nitrogen fixation; Transcription regulation; Activator;
KW ATP-binding; DNA-binding.
FT DOMAIN 33 193 A DOMAIN.
FT NP_BIND 219 448 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
FT NP_BIND 247 254 ATP (POTENTIAL).
FT NP_BIND 310 319 ATP (POTENTIAL).
FT DNA_BIND 501 520 H-T-H MOTIF.
SQ SEQUENCE 533 AA; 59546 MW; 0DEFBE02A1CF7695 CRC64;

Query Match 6.6%; Score 100; DB 1; Length 533;
Best Local Similarity 20.6%; Pred. No. 2.1;
Matches 81; Conservative 51; Mismatches 96; Indels 166; Gaps 19;

Qy 3 TEPERKGVVVGVGRAGSV-----RMRDLRNPSPSAFLNLIGF 42
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 88 TEEKKRGIVAVGEGITGVETGKAVARRLQHPDFLGRTRVSRNGKAKAAPFCVPIM 147
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 43 VSRELSDIGVQOISLEDALSSQEVAVICS-----ESSHEDYIRQF 87
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 RAQKVLGTI-AAERVYMPRLKKODVELLTMTATMIAPVELYLIENIERVLENENRRL 206
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 88 LNAGKHVLVE-----YPMTLSLAAQELWEL---AEQKGVLHHEHVVELLMEE 132
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 ----KHALKERFPSNLIIGNSKPM-----QEVYLIHKVASTKATVL----- 244
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 133 FAFLKKEVGKDLKLGSLFTA-----GPL-----EEERGFPAFGISRLTW 175
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 --ILGESGVGKELVANAIIHYNPNAAEALVTSNCAPLPENLAESLFGHEKSGFTGALT 302
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 176 LVSLF-----GELS-----LVSAT--- 189
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 HKGCFEQADGGTIFLDEVGELSPVQAKLVRLQNRTERVGGSKPVKVDVRIIAATNRN 362
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 190 ----LEER--KEDQYMKMTCLET-----EKKSP-----SWIEKGGPKLKR--- 225
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 LVEMVEQGTREDLYRLNVFPITVPPLRERGSVDIALADHFVSAFSRENGKNVKRISTP 422
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 226 --NRYLSHFHFKSGSLENVNVGVNKNIFLKDQNI 257
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 423 ALNMLMSYHW-PGNVRELENV-MERAVILSDDDV 454
: : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: November 1, 2002, 14:47:44
Job time : 17.2283 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:58 ; Search time 31.4084 Seconds
(without alignments)
905.569 Million cell updates/sec

Title: US-09-606-129A-3

Perfect score: 1508

Sequence: 1 MNTPEPKFGVVGVVGVGRAG.....RILHCLGLAEIQKYCCSRK 296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1508	100.0	296	2 G02066	biliverdin reducta
2	1490	98.8	296	2 S62624	biliverdin reducta
3	1247.5	82.7	295	2 A42268	biliverdin reducta
4	222.5	14.8	322	2 AF2324	biliverdin reducta
5	217	14.4	328	2 S74645	biliverdin reducta
6	157.5	10.4	371	2 F90486	dehydrogenase, pro
7	151.5	10.0	345	2 C65140	hypothetical 38.8
8	130.5	10.0	345	2 AH0994	probable oxidoredu
9	147.5	9.8	345	2 A91165	probable regulator
10	147.5	9.8	345	2 A86011	probable regulator
11	143	9.5	393	2 D69803	hypothetical prote
12	141	9.4	318	1 H62916	3-chlorobenzoate-3
13	136.5	9.1	364	2 A71443	probable inositol
14	136.5	9.1	374	2 AE2337	hypothetical prote
15	134	8.9	356	2 AC2319	hypothetical prote
16	133	8.8	338	2 AE1119	oxidoreductase hom
17	132	8.8	360	2 B72359	lipopolysaccharide
18	130.5	8.7	368	2 T39296	Oxidoreductase - f
19	130.5	8.7	371	2 S75555	hypothetical prote
20	130	8.6	343	2 A11428	oxidoreductases ho
21	129.5	8.6	324	2 B6819	oxidoreductase ypj
22	124.5	8.3	369	2 T38428	hypothetical oxido
23	123.5	8.2	352	2 T42744	hypothetical prote
24	123.5	8.2	371	1 D71201	hypothetical prote
25	122.5	8.1	358	2 G70026	conserved hypotet
26	122	8.1	338	2 AH1479	oxidoreductase hom
27	121.5	8.1	323	2 AH1479	oxidoreductase hom
28	121.5	8.1	324	2 AE3650	oxidoreductase (EC
29	121.5	8.1	346	2 F89784	hypothetical prote

30 121 8.0 341 2 A75025
31 119.5 7.9 328 2 H70014
32 119 7.9 343 2 AH1802
33 118.5 7.9 308 2 AH2713
34 118.5 7.9 334 2 E97495
35 117.5 7.8 357 2 A86658
36 116.5 7.7 323 2 AF1109
37 115 7.6 358 2 AC1345
38 115 7.6 358 2 AF1715
39 114.5 7.6 330 2 D98236
40 114.5 7.6 330 2 AG3049
41 114 7.6 360 2 A83004
42 112.5 7.5 390 2 AF3425
43 111.5 7.4 315 2 E75096
44 111.5 7.4 349 2 AB1715
45 111.5 7.4 349 2 AG1344

ALIGNMENTS

RESULT 1

G02066

biliverdin reductase (EC 1.3.1.24) - human

N:Alternate names: biliverdin IX-alpha reductase

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998

C:Accession: G02066

R:Komuro, A.; Tobe, T.; Nakano, Y.; Yamaguchi, T.; Tomita, M.

submitted to the EMBL Data Library, August 1995

A:Reference number: H00768

A:Accession: G02066

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-296 <COM>

A:Cross-references: EMBL:U34877; NID:g1143231; PID:g1143232

C:Keywords: oxidoreductase

Query Match 100.0%; Score 1508; DB 2; Length 296;

Best Local Similarity 100.0%; Pred. NO. 1.le-109;

Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNTPEPKFGVVGVVGVGRAGSVRMRLRNPHPSAFLNLIGVSRRELGSIDGVQOQISLE	60
Db	1	MNTPEPKFGVVGVVGVGRAGSVRMRLRNPHPSAFLNLIGVSRRELGSIDGVQOQISLE	60
Qy	61	DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTSLAAQELWELAEQKGV	120
Db	61	DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTSLAAQELWELAEQKGV	120
Qy	121	LHEEHVELLMEEFAFLKKEVVGKDLLKGSLLFTAGPLEEERFGPAPFSGISRLTWLVS	180
Db	121	LHEEHVELLMEEFAFLKKEVVGKDLLKGSLLFTAGPLEEERFGPAPFSGISRLTWLVS	180
Qy	181	GELSVATLEEREDQYMKMTVCLETCKSPSLWIEKGPGLKRNRYLSLPHFKSGSLEN	240
Db	181	GELSVATLEEREDQYMKMTVCLETCKSPSLWIEKGPGLKRNRYLSLPHFKSGSLEN	240
Qy	241	VPNVGVNKNIFLKQNTIFVQKLLQGFSEKELAAEKRLHCLGLAEETQKYCCSRK	296
Db	241	VPNVGVNKNIFLKQNTIFVQKLLQGFSEKELAAEKRLHCLGLAEETQKYCCSRK	296

RESULT 2

S62624

biliverdin reductase (EC 1.3.1.24) - human

N:Alternate names: biliverdin IX-alpha reductase

C:Species: Homo sapiens (man)

C:Date: 28-Oct-1996 #sequence_revision 09-May-1997 #text_change 21-Jul-2000

C:Accession: S62624; S62622; S29736

R:Maines, M.D.; Polevoda, B.V.; Huang, T.J.; McCoubrey Jr., W.K.

Eur. J. Biochem. 235, 372-381, 1996

A:Title: Human biliverdin IX-alpha reductase is a zinc-metalloprotein. Characterizati

A:Reference number: S62622; MUID:96202961

A:Accession: S62624

A:Molecule type: mRNA

A:Residues: 1-296 <MAT>

A:Cross-references: EMBL:X93086; NID:g1246748; PIDN:CAA63635.1; PID:g1246749

A:Accession: S62622

A:Molecule type: protein

A:Residues: 3-24,'X',26-27,'X',29-36;48-74;228-234;235-248 <NAF>

R:Maines, M.D.; Trakshel, G.M.

Arch. Biochem. Biophys. 300, 320-326, 1993

A:Title: Purification and characterization of human biliverdin reductase.

A:Reference number: S29736; MUID:9314333

A:Accession: S29736

A:Molecule type: protein

A:Residues: 3-24,'X',26-27,'X',29-36;48-74;228-234;235-248 <NAW>

A:Note: the sequence of peptide 1 from page 323 seems not to belong to this protein

C:Genetics:

A:Gene: BVR

C:Keywords: oxidoreductase

F:3-296/Product: biliverdin reductase IX-alpha #status experimental <MAT>

Query Match 98.8%; Score 1490; DB 2: Length 296;

Best Local Similarity 98.6%; Pred. No. 2.7e-108;

Matches 292; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MNTPEPKFGVVVGVGRAGSVRMRLRNPSPSAFLNLIGFVSRRRLGSDGVQOISLE 60

Db 1 MNAEPERKFGVVVGVGRAGSVRMRLRNPSPSAFLNLIGFVSRRRLGSDGVQOISLE 60

Qy 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120

Db 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120

Qy 121 LHERHVELLMEEFAFLKEVVGKDLLGSLFTAGPLEEERFGFPAPSGISRLTWLSLF 180

Db 121 LHERHVELLMEEFAFLKEVVGKDLLGSLFTAGPLEEERFGFPAPSGISRLTWLSLF 180

Qy 181 GELSIVSATLEERKEDQYMKMTCVLETKKSPLSWIEKGPGLKRNRYLSFHKSGSLEN 240

Db 181 GELSIVSATLEERKEDQYMKMTCVLETKKSPLSWIEKGPGLKRNRYLSFHKSGSLEN 240

Qy 241 VPNGVGNKNIPLKDNIFVQKLLGFSEKELAAEKKRILHCLGLAEIEQKYCCSRK 296

Db 241 VPNGVGNKNIPLKDNIFVQKLLGFSEKELAAEKKRILHCLGLAEIEQKYCCSRK 296

RESULT 3

A42268

biliverdin reductase (EC 1.3.1.24) - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A42268

R:Fakhrai, H.; Maines, M.D.

J. Biol. Chem. 267, 4023-4029, 1992

A:Title: Expression and characterization of a cDNA for rat kidney biliverdin reductase.

A:Reference number: A42268; MUID:92156147

A:Accession: A42268

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid; protein

A:Residues: 1-295 <FAK>

A:Cross-references: GB:M81681; NID:g203177; PIDN:AAA040830.1; PID:g203178

A:Experimental source: Kidney

A:Note: sequence extracted from NCBI backbone (NCBIP:82800)

C:Keywords: liver; oxidoreductase

Query Match 82.7%; Score 1247.5; DB 2: Length 295;

Best Local Similarity 82.6%; Pred. No. 1.8e-89;

Matches 245; Conservative 26; Mismatches 24; Indels 1; Gaps 1;

Qy 1 MNTPEPKFGVVVGVGRAGSVRMRLRNPSPSAFLNLIGFVSRRRLGSDGVQOISLE 60

Db 1 MDAEPKRFKFGVVVGVGRAGSVRLRLDKDPR-SAAFLNLIGFVSRRRLGSDGVQOISLE 59

Qy 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQKGV 120

Db 60 DALRSQETDVAYICSESSSHEDYIRQFLQAGKHVLYEYPMTLSFAAAQELWELAAQGRV 119

Qy 121 LHERHVELLMEEFAFLKEVVGKDLLGSLFTAGPLEEERFGFPAPSGISRLTWLSLF 180

Db 120 LHERHVELLMEEFEFLRVLGKELLASFTASPLEEERFGFPAPSGISRLTWLSLF 179

Qy 181 GELSIVSATLEERKEDQYMKMTCVLETKKSPLSWIEKGPGLKRNRYLSFHKSGSLEN 240

Db 180 GELSIVSATLEERKEDQYMKMTCVLETKKSPLSWIEKGPGLKRNRYVNFQFTSGSLEE 239

Qy 241 VPNGVGNKNIPLKDNIFVQKLLGFSEKELAAEKKRILHCLGLAEIEQKYCCSRK 296

Db 240 VPSGVGNKNIPLKDDIFVQKLLQVSAEDLAAEKKRIMHCLGLASDIQKLCHQK 295

RESULT 4

AF2324

biliverdin reductase [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

C:Accession: AF2324

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2324

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-322 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA075848.1; PID:g17133284; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: bvdR

Query Match 14.8%; Score 222.5; DB 2: Length 322;

Best Local Similarity 26.0%; Pred. No. 8.3e-10;

Matches 87; Conservative 52; Mismatches 110; Indels 85; Gaps 15;

Qy 13 VVGVRAGSVRMRLRNPSPSAFL-----NLIG-----FVSRRELGSIDGVQO 56

Db 3 IVGTGYAAKLK-----AEFLDESHLVAIAGSKLERTOTFAQAYQAEVITGWQO 53

Qy 57 ISLEDALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMTLSLAAQELWELAEQ 116

Db 54 L-----VEREDIDLVICTINRQHGAIARAALTAGKHVIVVEYPLSVDLTEAEELIALAKT 108

Qy 117 KGKVLHEHVELLMEEFAFLKEVVGKDLLGSLFF-----TAGP-----LEERF 162

Db 109 QOKLLHVEHIELLGLHQAALK-----QNLKVGHLFVRYSTINPQNPAPRKWTYNHAMF 163

Qy 163 GFPAFSGISRLTWLSLFGELISVATLEERK-----EQYMKMTVCLETERKSPLSW 215

Db 164 GFPLIGALSRLRLTDLFGKVFVNC--HORYWEIEEYVTCFCM-----TQLCFTSGL 216

Qy 216 -----IEKGPGL-KRNRYLSFHKSGSL-----ENVP-NVGWNKNIFLKQDN 256

Db 217 LAQVIYKGESLWQPERKFEVHGDNALIFDGTGFFIQSGESTPVELTTRRGLFAKDT 276

Qy 257 IFVQKLLGQFSEKELAAEKKRILHCLGLAEIEQK 290

Db 277 MWLDHI---FDGAPLYVTPPEESLYTLKVAQAOR 307

RESULT 5

S74645

biliverdin reductase (EC 1.3.1.24) - Synechocystis sp. (strain PCC 6803)

N:Alternate names: hypothetical protein sir1784

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S74645

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S74645

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-328 <RAN>

A:CROSS-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAAL16797.1; PID:d101753

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Keywords: oxidoreductase

Query Match 14.4%; Score 217; DB 2; Length 328;
Best Local Similarity 24.9%; Pred. No. 2.3e-09;

Matches 79; Conservative 57; Mismatches 139; Indels 42; Gaps 11;

Qy 11 VVVGVRAGSVRMRLRNPHPPSAFLNLIGFVSRRRLGSID-----GVQ-QISLEDAIS 64

Db 13 VGI VGTGYAQRRAEVRFGDRRS-----OLVSFWGNSEANTAKFADTFGVRPQQSQWQALIN 68

Qy 65 SQEVEVAYICSESSSHEDYIRQFLNAGKHVLYVEPMTLSLAAQELWELABQKGVLHEE 124

Db 69 DPEIDLVLTIATINLHGAI EAALQAGKHVVLYEPLALTYAMGKKLOOLAREKGLLHVE 128

Qy 125 HVELLMEEFAPLKVEV-VGKDL-LKGSLLFTAGPLBE-----ERFGFPAFSISRLTW 175

Db 129 HIELGGVHOAIRONLKIGEVFYARYSTINGQNPAPQRWTHHQQEGFPLVAALSRSR 188

Qy 176 LVSLFGELSLVSATLE--ERKEDQYMK--MTVCLETEKKSPLSWIEEKGGLKRN-RYL 229

Db 189 FTDLFGTVQVQDAOCRFWDQNPFPYFRACLATATAYLQFNNGLKAEVYICKGVFVHONERIF 248

Qy 230 SFHFKSGSL-----ENVPNVGVNKNIFLKDQNIQFVKLLGQFSEKELAAEK 275

Db 249 TLHGDRGLTFVGETGLIQGTETITVGSRRGLFRQDTEAVLDYLT---TGRPLYVDL 305

Qy 276 KRIHLGLAEIKQYC 292

Db 306 EASLYALEVADLCAQAC 322

RESULT 6

F90486

dehydrogenase, probable [imported] - *Sulfolobus solfataricus*

C:Species: *Sulfolobus solfataricus*

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: *Sulfolobus solfataricus* complete genome.

A:Reference number: A99139

A:Accession: F90486

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-371 <KUR>

A:CROSS-references: GB:AE006641; NID:g13816451; PIDN:AAK43149.1; GSPDB:GN00155

C:Genetics:

A:Gene: SS03049

Query Match 10.4%; Score 157.5; DB 2; Length 371;
Best Local Similarity 19.9%; Pred. No. 0.00011;

Matches 64; Conservative 59; Mismatches 118; Indels 81; Gaps 10;

Qy 7 RKFGVVVGVGRAGSVRMRLRNPHPPSAFLNLIGFVSR-----RELGSIDGV-QQISLE 60

Db 18 KRIGVAVVGLSGIKGTHVKALKDLEKETEFVKLVAVVDQIKAIKIEKSGYGTPTTTID 77

Qy 61 DALSSQVEVAYICSESSSHEDYIRQFLNAGKHVLYVEPMTLSLAAQELWELAEOK---- 117

Db 78 EVLNSEVDVVISIATPSYLAHAPAILAIEYKGVIVKPMATTLAGAREMVSRAERNEVK 137

Qy 118 -GKVLHEEH-----VELLMEEF-----AFLLKKEVVGKDLLKGSLLF 152

Db 138 LGVIFQERYAPDIRRLKNDILKGLRIVLIESELKWKYRDMKGYKRDIEARS----- 189

Qy 153 TAGPLEERERGFPAFSGISRLTWLVSFLGELSLVSATLEERKEKDOYMKMTVCLETEKSK 212

Db 190 WRGMNTEGGGVMTNQGLHTIDLMIWLNGEVEEVSGFVDNLTHDG-----IE----- 236

Qy 213 LSWIEEKGPGILKRNRYLSFHFKSGSLENVPNV-----GVNKNIFLKIDQN 256

Db 237 ---VEDTAVAIMR-----YKNCALGTISQTVSMKPTTYQYRKIRVNGSNGFVEITDGS 286

Qy 257 IFVOKLLGQFSEKELAAEKRR 277

Db 287 LSTVAIEGKIEESKSSVEYKK 307

RESULT 7

C65140

hypothetical 38.8 kD protein in gntR-ggt intergenic region - *Escherichia coli* (strain

C:Species: *Escherichia coli*

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999

C:Accession: C65140

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: C65140

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-345 <BLAT>

A:CROSS-references: GB:AE000420; GB:U00096; NID:g1789840; PIDN:AAC76465.1; PID:g17898

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yhhx

C:Superfamily: conserved hypothetical protein bi624

Query Match 10.0%; Score 151.5; DB 2; Length 345;
Best Local Similarity 31.3%; Pred. No. 0.0003;

Matches 42; Conservative 27; Mismatches 52; Indels 13; Gaps 4;

Qy 59 LEDALSQSEVAVICSESSSHEDYIRQFLNAGKHVLYVEPMTLSLAAQELWELAEOKG 118

Db 57 LDEVLDNDPDVCLVVVCTHADSHFEYAKRALEAGKNVLVEKPTPTTLAQAKELFALAKSG 116

Qy 119 KVLHEEHVELLMEEFAPLKKEVVGKDLLKGSLLFTAG-----PLEERFGFP---AP-- 167

Db 117 LTVTPYNNRRDFSCFLTAKKAIESGKL--GEIVEVESHFYRVPVETKPGLPDQAGAFYG 174

Qy 168 SGISRLTWLSLFG 181

Db 175 LGVHTMDQIISLFG 198

RESULT 8

AH0994

probable oxidoreductase STV4266 [imported] - *Salmonella enterica* subsp. *enterica* sero

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi

A>Note: this species has also been called *Salmonella typhi*

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001

C:Accession: AH0994

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

, S.; Mouton, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A>Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se

A:Reference number: AB0502; PMID:11677608

Qy 180 FGEISLVSATLEERKEDQYKMTVCLETETKSKPLSWI--EEKGPG 222
 | : : || | |
Db 166 RADTS-----RPNWYSSKEKGG 184

RESULT 15
AC2319
hypothetical protein all4106 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC2319
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriouchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2319
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-356 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075805.1; PID:g17133241; GSFDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all4106

Query Match 8.9% Score 134; DB 2; Length 356;
Best Local Similarity 24.9%; Pred. No. 0.0071;
Matches 44; Conservative 28; Mismatches 53; Indels 52; Gaps 5;

Qy 11 VVVVGVRGAGSVRMRD-----LRNHPSSAF-----NLIGFVSRRRLGS 50
 : : || | | : : |
Db 4 IAVIGVGRGWVHLRNFLHPQAEVVAIVDPHPERLAVVQKFKAENVLLTTDWSDLQK 63

Qy 51 IDGVQOISLEDALSSQEEVEVAYICSESSSHEDYIRQFLNAGKHVLYEYPMFTLSLAAQEL 110
 | : : | : | | | | | |
Db 64 VPGLTAVA-----IATPATTHYALIKDALAOGYHVLAEKPLTLDPEVCEQL 109

Qy 111 WELAEQKGVLEHREHVELL---MEEFAPLKKEVVGKDLLKGSLLFTAGPLEEERFGF 164
 : || : | | | | | | |
Db 110 QLAEQRLILMVDHTYLFHPAVEE-----GQTVVQAGKIGELRIGY 151

Search completed: November 1, 2002, 14:50:55
Job time : 32.575 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:44:48 ; Search time 24,746 Seconds
(without alignments)
292.168 Million cell updates/sec

Title: us-09-606-129a-3
Perfect score: 1508
Sequence: 1 MNTEPRKGVVGVGVGRAG.....RILHCLGLAEIQKYCCSRK 296

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCRUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123.5	8.2	315	4	US-09-720-817-3
2	123.5	8.2	315	4	US-09-720-817-6
3	122.5	8.1	280	4	US-09-720-817-1
4	112	7.4	229	4	US-08-936-165A-519
5	112	7.4	332	4	US-09-036-987A-15
6	112	7.4	332	4	US-09-370-700-15
7	97	6.4	988	2	US-08-286-819A-19
8	97	6.4	988	3	US-08-980-357-19
9	92.5	6.1	351	1	US-08-402-217A-2
10	92.5	6.1	351	1	US-08-700-178-2
11	92.5	6.1	351	3	US-08-995-654-2
12	91	6.0	264	2	US-07-857-224B-9
13	90	6.0	551	3	US-08-886-886-15
14	89	5.9	759	1	US-08-676-967-1
15	89	5.9	759	1	US-08-676-974-1
16	89	5.9	759	2	US-09-098-487-1
17	86.5	5.7	1199	4	US-09-208-742-2
18	86.5	5.7	1199	4	US-09-332-295-4
19	85.5	5.7	514	2	US-08-960-022-14
20	84.5	5.6	487	2	US-08-712-709-8
21	84.5	5.6	487	3	US-09-111-444-8
22	84.5	5.6	487	4	US-09-541-228-8
23	84	5.6	1205	4	US-09-330-330-1
24	83	5.5	573	2	US-08-745-934-4
25	83	5.5	573	4	US-09-147-009-7
26	82.5	5.5	270	2	US-08-852-743-5
27	82.5	5.5	270	3	US-09-185-370-5

ALIGNMENTS

RESULT 1
US-09-720-817-3
; Sequence 3, Application US/09720817
; Patent No. 6340579
; GENERAL INFORMATION:
; APPLICANT: Masai, Ei-ji
; APPLICANT: Fukuda, Masao
; APPLICANT: Katayama, Yoshihiro
; APPLICANT: Nishikawa, Seiji
; APPLICANT: Hotta, Yasushi
; TITLE OF INVENTION: NOVEL GENE AND TRANSFORMANT CARRYING THE SAME
; FILE REFERENCE: 200966050XPCT
; CURRENT APPLICATION NUMBER: US/09/720,817
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: JAPAN 10/203200
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PCT/JP99/03410
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Sphingomonas sp. CR-0310201
US-09-720-817-3

Query Match	8.2%	Score 123.5;	DB 4;	Length 315;
Best Local Similarity	20.0%	Pred. No. 6.9e-05;		
Mismatches	66;	Conservative	53;	Mismatches 136;
Indels	75;	Gaps	12;	
Qy	11	VVVGVGRAGSRMRDLRNPSPSAFLNLGIFVSR-----ELGSDGVQVQ--ISLEDAL 63		
Db	3	IALGAGAFGEKHLGDKLNDG-----VEIVSIIRKAEQAEEAAVAAKYGAKHSGTDLSEAL 58		
Qy	64	SSQVEVAYTCSESSSESHEDYIRQFLNAGKHLVPEYPTLSLAAQELWELAEQKGVLHE 123		
Db	59	ARDDVDVAVILCTPTQMHAEQAIAACNAGKHVQVEIPLADSWADAEAVMKKSOETGLVCMV 118		
Qy	124	EHVELLMEEFAFLKVEYGVGDKLKGSLFTAGLPLEERFEGFPF-----SG 169		
Db	119	GHTRRFNPSHQYIHNKIV-----AGELAIQMDVQTYFFRRKNNNAKGEPRSW 166		
Qy	170	ISRLTW-----LVSLFG-----ELSVSATLEERKEDQYMKMTVCLE 206		
Db	167	TDLLHWHHAHTVDLFAYQAGKIYQANAVQGPPIHPELGIAIMDSIOLKSETGAICTLSLS 226		
Qy	207	TEKSPSL-----WTEKGPGLKRNRYLSFHFKSGSLE--NVPNGVKNKN-IFLKQDNIFV 259		
Db	227	FNDGPLGTFRFYICDNGTWIARYDDL-----VTGKEPVDVSKVDVSMNGIELQDREFIA 282		

TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 519:
SEQUENCE CHARACTERISTICS:
LENGTH: 229 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-936-165A-519

Query Match 7.4%; Score 112; DB 4; Length 229;
Best Local Similarity 22.8%; Pred. No. 0.0007;
Matches 39; Conservative 30; Mismatches 70; Indels 32; Gaps 5;
Qy 63 LSQVEVEVAYICSSSHSHEDYIRFLNAGKHVLVEYPMFLSLAAQELWELAEQKGVHL 122
Db 54 LKDDTIDVHVCTPNDSHCEITVAGLHAGKHVMCEKPMKTTAEQAQKMTIDTAKSTGKKLT 113
Qy 123 EHVLLMEEFALFKKEVGVGKDLKGSLLF-----TAGPL--EEERGFPAFS 168
Db 114 IGYNRRFPDSQFLHQAQRGL--GDIYFGKAHAIRRAVPTWGVFLNEEAQGGGLID 171
Qy 169 ----GISRLTWLVSFGELSLVSATLEERKEDQYMKMTVCLETKKSPLSW 215
Db 172 IGTHALDLTLWMDNYEPESVMGSTFHK-----LNKQHDAPNAW 210

RESULT 5
US-09-036-987A-15
Sequence 15, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036.987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 332 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-036-987A-15

Query Match 7.4%; Score 112; DB 4; Length 332;
Best Local Similarity 24.5%; Pred. No. 0.0012;
Matches 47; Conservative 34; Mismatches 69; Indels 42; Gaps 7;
Qy 42 FVSRRELGSIDGVOOISLEDALSSQVEVEVAYICSSSHSHEDYIROFLNAGKHVLVEYPM 101
Db 46 FAARFECEAVLGQYRL-----LERPDIDAVYVPLPPGHHAEWIGKALEADKHVLAERPLT 100
Qy 102 LSLAAQELWELAEQKGVHLHEEHVELLMEEFAFLKKEVGVGKDLKGSLLFTAGPLEEER 161
Db 101 TTASDTARLVGLARRKNLLRENVL-----FLHGRHDVVVDLLQ-----SCEIGELR 148
Qy 162 -----FGPPAF--SGISRLTWLVS-----LFGELSLVSATLEERKEDQ 197
Db 149 EFTAVFGIPPLPDTDIRYRTELGGGALLDIGVYPARAARHFLGLPLTVLGASSHEAQESG 208
Qy 198 Y-MKMTVCLETE 208
Db 209 VDLSGSVLLQSE 220

RESULT 6
US-09-370-700-15
Sequence 15, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 332
TYPE: PRT
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-15

Query Match 7.4%; Score 112; DB 4; Length 332;
Best Local Similarity 24.5%; Pred. No. 0.0012;
Matches 47; Conservative 34; Mismatches 69; Indels 42; Gaps 7;
Qy 42 FVSRRELGSIDGVOOISLEDALSSQVEVEVAYICSSSHSHEDYIROFLNAGKHVLVEYPM 101
Db 46 FAARFECEAVLGQYRL-----LERPDIDAVYVPLPPGHHAEWIGKALEADKHVLAERPLT 100
Qy 102 LSLAAQELWELAEQKGVHLHEEHVELLMEEFAFLKKEVGVGKDLKGSLLFTAGPLEEER 161
Db 101 TTASDTARLVGLARRKNLLRENVL-----FLHGRHDVVVDLLQ-----SCEIGELR 148
Qy 162 -----FGPPAF--SGISRLTWLVS-----LFGELSLVSATLEERKEDQ 197
Db 149 EFTAVFGIPPLPDTDIRYRTELGGGALLDIGVYPARAARHFLGLPLTVLGASSHEAQESG 208
Qy 198 Y-MKMTVCLETE 208
Db 209 VDLSGSVLLQSE 220

RESULT 7
US-08-286-819A-19


```

US-08-402-217A-2
; TOPOLOGY: linear
; MOLECULE TYPE: protein
Query Match      6.4%; Score 97; DB 3; Length 988;
Best Local Similarity 20.8%; Pred. No. 0.25;
Matches 66; Conservative 49; Mismatches 106; Indels 96; Gaps 14;

QY   21 SVRMRLRNP-----HPSSAFNLNIGFVSRRELIGSDGVQQISLEDALSSOEVEAYICS 75
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   55 AVCLAVLRYPGWYTHIKSPDSVIQISKQ-----IGVSPSSL-DHYPQRENTLWDHLK 108
QY   76 ESSSHEDYIROFLNAGKHVLVEYPWTLSLAAQAQLWELAEQGKGKVULHEHVELLMEEFPAP 135
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   109 EIRSEYDFV-----TFTILSEYRTMF-----KYHLQALANGDATH-----LLHCIDIF 151
QY   136 LKEVVVGKDLKGLSLLFTAGPLEEERFGPPAFSGISRITLWLSVLFGEISLVSATLEERKE 195
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   152 LRK-----NKIILPAITTLERMVMWEARMAEKKLFNIVSVKSLTN 190
QY   196 DOYMKMTVCLETE-----KKSPLSWIEEEKGPG-----LKNRYLSFHFKSGSLEN 240
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   191 EQKELEGITTSQHPSESNTILGWLKEP-PCHPSPETFLKIIERLEYI-----RGMDLET 245
QY   241 VPMVGNKNIFLK-----DQN-----IFVKLLGGQFSEKELAIAEKKR 277
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   246 VOISHLRNRLLQSRLGRSYEPYAFRFQENKRYSLTIYLTLTQELTDRAFEIHDRQ 305
QY   278 ILHCLG-----LAEEIOK 290
    || | | | | | |
Db   306 ILSLSKGRKAQEIOK 322

RESULT 9
US-08-402-217A-2
; Sequence 2, Application US/08402217A
; Patent No. 5587301
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Sellhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,217A
; FILING DATE: 10-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0028US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-980-357-19
; TOPOLOGY: linear
; MOLECULE TYPE: protein
Query Match      6.1%; Score 92.5; DB 1; Length 351;
Best Local Similarity 25.1%; Pred. No. 0.15;
Matches 72; Conservative 39; Mismatches 85; Indels 91; Gaps

QY   54 VOQISLE-DALSSOEVEVA-----YICS-----ESSSHEDYIROFLNAGKHVLVEYPM 100
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   18 LOOKLOLDSLQOQEKLSSSHOKLSFOEQEMAKEKNLFEELKOTLD----- 66
QY   101 TLSLAAAOELWELAQKGVULHEHVHVELLMEEFAFLKKEVVCK--DLLKGSLLFT-AGPL 157
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   67 --ELDKLOQKEBAQRVLKQL-EDEAKSRAEBELKLEBKLGKGAELEKSSNAHTQTALL 123
QY   158 EBEREG-----FPAFSGISRITLWLSVLFGEISLVSATLEER----- 193
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   124 LEKYDYSMVQSLEDVTAQPESYKALT--ASETEDLKLENSLSQEKVAKAGNAEDVHQOI 181
QY   194 -----KEOYMKMTVCLETEKKSPLSWIEEEKGPGUKRNRYLSFHFKSGSLENVPMVGNK 248
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   182 LATESSNOEYVRMLDLOLT--KSALKETEIK-----ETTVSFLQKITDLOQO----- 226
QY   249 NIFLDQNIFFVKLL-----GFSEKELAIAEKKRIULHCLGLAEIQKY 291
    :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   227 ---LKOQBEDFRKQLEDEGRKAEREKNTAE-----LTEEINKWK 262

RESULT 10
US-08-700-178-2
; Sequence 2, Application US/08700178
; Patent No. 5783669
; Patent No. 5783669 5700912
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Sellhamer, Jeffrey J.
; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
; TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,178
; FILING DATE: August 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,217
; FILING DATE: March 10, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0028-1 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-700-178-2

```


Qy 135 -FLKKEVVGKOLLKGSLLFTAGPLLEERFGPPAFSGISRLTWLVSFLGELSLVSATLEER 193
Db 382 QFMTQEAQAOKCLLAASPENEAGGL-----K 406
Qy 194 KEDQYMKMTVCLETEKKSPLSWEIEKGPLKRNRYLSFH--FKSG--SLENVPNVGVN-- 247
Db 407 LDGROLKVDLAVTRDEAAKLQTTKVKRPTGTRNLYLAREGLIRAGTKAAEGVSAADMAKR 466
Qy 248 -----KNIFLKQDNIFVQK 261
Db 467 ERFELLKHQKLKQDNIFVSR 486

RESULT 15
US-08-676-974-1
; Sequence 1, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 759 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-676-974-1

Query Match 5.9%; Score 89; DB 1; Length 759;
Best Local Similarity 21.5%; Pred. No. 1.2;
Matches 56; Conservative 44; Mismatches 100; Indels 60; Gaps 9;

Qy 22 VRMRDLRNPFPSSAFNLICFVSRRELGSIDGVQOISLEDAALSSQEEVEVAICSESSSHE 81
Db 267 IOKRAVRPAPAKSDHSEEDSDLEESDSDIDGEELAQSDTSTEEQEDKAVQVSNKKRKR 326
Qy 82 DYIRQFLNAGKHVLVEYPMTLISL-AAQAELMELAEQKQK-----VLHEEHVELLMEEFA 134
Db 327 --LPSDVNEGKTVFIR--NLSFDSSEELGELLQFQGLKYYRIVLHPDTEHSGKCAFA 381
Qy 135 -FLKKEVVGKOLLKGSLLFTAGPLLEERFGPPAFSGISRLTWLVSFLGELSLVSATLEER 193
Db 382 QFMTQEAQAOKCLLAASPENEAGGL-----K 406
Qy 194 KEDQYMKMTVCLETEKKSPLSWEIEKGPLKRNRYLSFH--FKSG--SLENVPNVGVN-- 247
Db 407 LDGROLKVDLAVTRDEAAKLQTTKVKRPTGTRNLYLAREGLIRAGTKAAEGVSAADMAKR 466

Qy 248 -----KNIFLKQDNIFVQK 261
Db 467 ERFELLKHQKLKQDNIFVSR 486
Search completed: November 1, 2002, 14:51:55
Job time : 25.746 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:40:53 ; Search time 67.0997 Seconds
(without alignments)
489.985 Million cell updates/sec

Title: US-09-606-129A-3

Perfect score: 1508

Sequence: 1 MNTPEPKFGVVVGVGRAG.....RILHCLGLAEIQKYCCSRK 296

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225	14.9	60	21	AAG02024 Human secreted pro
2	182	12.1	63	21	AAG02023 Human secreted pro
3	151.5	10.0	434	22	ABG21584 Novel human diagno
4	147	9.7	241	22	ABG21582 Novel human diagno
5	147	9.7	389	22	ABG29497 Novel human diagno
6	130.5	8.7	335	22	ABG60024 Drosophila melanog
7	124.5	8.3	226	22	AAU49252 Propionibacterium
8	123.5	8.2	175	18	AAW28328 Amino acid sequenc
9	123.5	8.2	315	21	AAV68460 Sphingomonas sp. C
10	122.5	8.1	280	21	AAV68459 Sphingomonas sp. C
11	121	8.0	341	22	AAB96416 Putative P. abyssi

12	119.5	7.9	346	22	AAU36865 Staphylococcus aur
13	116.5	7.7	337	22	AAG92054 C glutamicum prote
14	116.5	7.7	337	22	AAB79386 Corynebacterium g1
15	116.5	7.7	341	22	AAU34111 Staphylococcus aur
16	112	7.4	229	19	AAW77759 Glucose-fructose o
17	112	7.4	332	20	AAV39310 SpnN protein invol
18	112	7.4	336	22	AAB70956 S. spinosa protein
19	112	7.4	344	22	AAU04826 Micromonospora eve
20	112	7.4	357	22	AAU34110 Staphylococcus aur
21	112	7.4	359	22	AAU36864 Staphylococcus aur
22	111.5	7.4	315	22	AAB96316 Putative P. abyssi
23	110.5	7.3	327	22	AAU03634 Group B Streptococ
24	105	7.0	329	22	AAE00480 Streptomyces galil
25	102.5	6.8	702	22	AAU49114 Propionibacterium
26	100	6.6	307	22	AAG98892 E. coli growth and
27	99.5	6.6	413	22	AAU56511 Propionibacterium
28	98.5	6.5	1316	22	AAG81126 Mycobacterium tube
29	98	6.5	1374	22	AAB69070 Human male enhance
30	97	6.4	359	22	AAG93130 C glutamicum prote
31	97	6.4	359	22	AAB79395 Corynebacterium g1
32	96	6.4	988	13	AAR24302 Transposase from g
33	95.5	6.3	344	22	AAU43141 Propionibacterium
34	95	6.3	219	19	AAW98401 H. pylori GPO 138
35	95	6.3	753	22	ABG20518 Novel human diagno
36	94.5	6.3	360	21	AAG23569 Arabidopsis thalia
37	94.5	6.3	363	21	AAG23568 Arabidopsis thalia
38	94.5	6.3	367	21	AAG23567 Arabidopsis thalia
39	94.5	6.3	463	21	AAG42560 Arabidopsis thalia
40	94.5	6.3	467	21	AAG42559 Arabidopsis thalia
41	94.5	6.3	506	21	AAG42558 Arabidopsis thalia
42	93.5	6.2	867	22	AAB96297 Putative P. abyssi
43	93	6.2	341	22	AAG89934 C glutamicum prote
44	93	6.2	341	22	AAB79394 Corynebacterium g1
45	93	6.2	417	22	ABB67972 Drosophila melanog

ALIGNMENTS

RESULT 1

AAG02024

ID AAG02024 standard; Protein: 60 AA.

AC AAG02024;

XX 06-OCT-2000 (first entry)

DT Human secreted protein, SEQ ID NO: 6105.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

DE gene therapy; chromosome mapping.

KW Homo sapiens.

OS EPI033401-A2.

PN 06-SEP-2000.

PD 21-FEB-2000; 2000EP-0200610.

PF 26-FEB-1999; 99US-0122487.

PR (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

PA WPI: 2000-500381/45.

PI N-PSDB; AAC02030.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

DR obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 13; SEQ ID 6105; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.

XX
SQ Sequence 60 AA;

Query Match 14.9%; Score 225; DB 21; Length 60;
Best Local Similarity 97.9%; Pred. No. 2.1e-14;
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 MRDLRNPSPSAFLNLGFGVSRRELSDGVDGQQISLEDAISSQVEV 70
|||||
DB 1 MRDLRNPSPSAFLNLGFGVSRRELSDGVDGQQISLEDAISSQVEV 47

RESULT 2

AAG02023

ID AAG02023 standard; Protein; 63 AA.

XX
AC AAG02023;

XX
DT 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 6104.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX
PD 06-SEP-2000.

XX
PF 21-FEB-2000; 2000EP-0200610.

XX
PR 26-FEB-1999; 99US-0122487.

XX
PA (GSEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX
PI WPI; 2000-500381/45.

DR N-PSDB; AAC02029.

XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX
PS Claim 13; SEQ ID 6104; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC

CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.

XX
SQ Sequence 63 AA;

Query Match 12.1%; Score 182; DB 21; Length 63;
Best Local Similarity 92.3%; Pred. No. 3.3e-10;
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNYEPERKFGVVVGVGRCAGSVRMRDLRNPSPSAFLNL 39
|||||
DB 1 MNAEPERKFGVVVGVGRCAGSVRMRDLRNPSPSAFLNL 39

RESULT 3

ABG21584

ID ABG21584 standard; Protein; 434 AA.

XX
AC ABG21584;

XX
DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #21575.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX
OS Homo sapiens.

XX WO200175067-A2.

XX
PD 11-OCT-2001.

XX
PF 30-MAR-2001; 2001WO-US08631.

XX
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX
PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX
PI WPI; 2001-639362/73.

DR N-PSDB; AAS85771.

XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX
PS Claim 20; SEQ ID No 51943; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO

C. glutamicum, identify and localise C. glutamicum sequences of interest, in evolutionary studies, in determining SMP protein regions required for function, in modulating SMP protein activity, in modulating the metabolism of sugars, and in modulating high-energy molecule production in a cell (i.e. ATP, NADPH).

Sequence	337 AA;
SQ	

Query Match 7.7%; Score 116.5; DB 22; Length 337;

Best Local Similarity 25.1%; Pred. No. 0.0076;
Matches 42; Conservative 24; Mismatches 70; Indels 31; Gaps 4;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Qy 11 VVVGVGRAGSVRMRDLRN-----PHPSSAFLNLIGFVSRRELGSIDGVQQIS 58

Db 5 LALICAGRIGSNHARLITHVIGSELVAVDPTPNAETL-----ADELGA---VAFSN 54

Qy 59 LEDALSSQEVEVAYICSESSSHEDYIRQFLNAGKHVLVEYPMTLSLAAQELWELAEQKG 118

Db 55 PDDVITRDDIDAVLIATPARTHADLVVKAAGKHVFVEKPMVLTEDADRINAAREAN 114

QY 119 KVLHEEHVELLMEEFAFLKKEVWGKDLLKGSLL-----FTAGP 156

Db 115 TVLQVGNNRFAAGFAAARARIDAGDIGTLPOLLRSVTRDPGPFADP 161

RESULT 15

AAU34111

ID AAU34111 standard; Protein; 341 AA.

AC AAU34111;

DT 13-FEB-2002 (first entry)

DE Staphylococcus aureus cellular proliferation protein #387.

KW Antisense; prokaryotic cellular proliferation protein;

KW antibiologic; antibacterial; drug design.

OS Staphylococcus aureus.

PN WO200170955-A2.

27-SEP-2001.

21-MAR-2001: 2001WO-US09180.

21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P;

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

DR WPI; 2001-611495/70.

DR N-PSDB; AAS51970.

New polynucleotides for the identification and development of antitoxics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 5607: 511bp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:08 ; Search time 1.2717 Seconds
(without alignments)
952.238 Million cell updates/sec

Title: US-09-606-129a-18

Perfect score: 41

Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:**
1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mhc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp_vertebrate:**
14: sp_unclassified:**
15: sp_rvirus:**
16: sp_bacteriap:**
17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	41	100.0	599	10 Q9FHV3	Q9fHV3 arabidopsis
2	41	100.0	633	10 Q49511	Q49511 arabidopsis
3	38	92.7	295	11 Q9CY64	Q9cy64 mus musculus
4	38	92.7	296	4 Q9BRW8	Q9brw8 homo sapien
5	38	92.7	296	4 Q96QL4	Q96ql4 homo sapien
6	38	92.7	303	11 Q9DD21	Q9dd21 mus musculus
7	37	90.2	106	5 Q9XZM5	Q9xzm5 leishmania
8	36	87.8	512	10 Q9LML0	Q9lml0 arabidopsis
9	35	85.4	508	11 Q9D6S2	Q9d6s2 mus musculus
10	34	82.9	279	11 Q9QXV3	Q9qxv3 mus musculus
11	34	82.9	401	10 Q9M149	Q9m149 arabidopsis
12	34	82.9	431	10 Q04613	Q04613 arabidopsis
13	33	80.5	189	2 Q03949	Q03949 anabaena sp
14	33	80.5	267	2 Q9AKS0	Q9aks0 pseudomonas
15	33	80.5	495	10 Q9ST63	Q9st63 solanum tub
16	33	80.5	499	10 Q941V1	Q941v1 oryza sativ

17	33	80.5	508	10 Q80874	Q80874 arabidopsis
18	32	78.0	102	10 Q9S738	Q9s738 lycopersico
19	32	78.0	104	15 Q9Q553	Q9q553 human immun
20	32	78.0	108	16 P75909	P75909 escherichia
21	32	78.0	111	4 Q9H007	Q9h007 homo sapien
22	32	78.0	142	4 Q9BTS7	Q9bts7 homo sapien
23	32	78.0	233	4 Q9UIJ2	Q9uij2 homo sapien
24	32	78.0	235	4 Q9P0U6	Q9p0u6 homo sapien
25	32	78.0	279	4 Q43658	Q43658 homo sapien
26	32	78.0	279	4 Q43658	Q43658 homo sapien
27	32	78.0	279	4 Q9HD98	Q9hd98 homo sapien
28	32	78.0	283	4 Q9UKS2	Q9uks2 homo sapien
29	32	78.0	294	4 Q9HD99	Q9hd99 homo sapien
30	32	78.0	294	4 Q00532	Q00532 homo sapien
31	32	78.0	338	3 Q9CIL3	Q9cil3 neurospora
32	32	78.0	377	16 Q9KSC1	Q9kscl vibrio chol
33	32	78.0	422	4 Q9UK53	Q9uks3 homo sapien
34	32	78.0	422	4 Q9UIJ4	Q9uij4 homo sapien
35	32	78.0	455	5 Q9VJ28	Q9vj28 drosophila
36	32	78.0	497	10 Q9C8V1	Q9c8v1 arabidopsis
37	32	78.0	516	5 Q21592	Q21592 caenorhabd
38	32	78.0	571	10 Q24022	Q24022 lycopersico
39	32	78.0	590	10 Q98S67	Q98s67 gullardia
40	32	78.0	601	4 Q9BRB5	Q9brb5 homo sapien
41	31	75.6	109	5 Q9U6W5	Q9u6w5 caenorhabd
42	31	75.6	137	4 Q9VZM4	Q9vzm4 drosophila
43	31	75.6	137	4 Q9NSR1	Q9nsr1 homo sapien
44	31	75.6	152	5 Q44797	Q44797 caenorhabd
45	31	75.6	153	14 Q91UN1	Q91un1 plasmid psb
			234	12 Q91GH4	Q91gh4 epiphyas po

ALIGNMENTS

RESULT 1

Q9FHV3 ID Q9FHV3 PRELIMINARY: PRT: 599 AA.
AC Q9FHV3;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE MUTATOR-LIKE TRANSPOSASE-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=93937451; PubMed=10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT pl and TAC clones."
RL DNA Res. 6:183-195(1999).
DR EMBL; AB017068; BAB1366.1; .
SQ SEQUENCE 599 AA; 69407 MW; 36BAEE2F2A82D717 CRC64;

Query Match 100.0%; Score 41; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 0.95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||||||
Db 563 KKRIMHC 569

RESULT 2

Q49511 ID Q49511 PRELIMINARY: PRT: 633 AA.
AC Q49511;

DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MUDR TRANSPOSABLE ELEMENT - LIKE PROTEIN (MUDR TRANSPOSABLE ELEMENT-
DE LIKE PROTEIN).
GN F26J12.70 OR AT4C18410.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021710; CAAL6721.1; -;
DR EMBL; AL161548; CAB78843.1; -;
DR InterPro: IPR004332; MUDR.
DR Pfam: PF03108; MUDR.1;
SQ SEQUENCE 633 AA; 72930 MW; 3FC298BF2218C623 CRC64;
Query Match 100.0%; Score 41; DB 10; Length 633;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KKRIMHC 7
Db 591 KKRIMHC 597
RESULT 3
Q9CY64 PRELIMINARY; PRT; 295 AA.
ID Q9CY64
AC Q9CY64;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE 2500001N03RIK PROTEIN.
GN 2500001N03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guncich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni J., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010847; BAB27219.1; -;
DR MGD; MGI:1917355; 2500001N03RIK.
DR InterPro: IPR000683; GFO_IDH_MoCa.
DR Pfam: PF01408; GFO_IDH_MoCa; 1.
SQ SEQUENCE 295 AA; 33524 MW; F2E1682BD77032A4 CRC64;
Query Match 92.7%; Score 38; DB 11; Length 295;
Best Local Similarity 85.7%; Pred. No. 2.2;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KKRIMHC 7
Db 274 KKRILHC 280
RESULT 4
Q9BRW8 PRELIMINARY; PRT; 296 AA.
ID Q9BRW8
AC Q9BRW8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SIMILAR TO BILIVERDIN REDUCTASE A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN, PRIMITIVE NEUROECTODERMAL;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005902; AAH05902.1; -;
DR InterPro: IPR000683; GFO_IDH_MoCa.
DR Pfam: PF01408; GFO_IDH_MoCa; 1.
SQ SEQUENCE 296 AA; 33432 MW; A389AFDDB6ADBABF CRC64;
Query Match 92.7%; Score 38; DB 4; Length 296;
Best Local Similarity 85.7%; Pred. No. 2.2;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KKRIMHC 7
Db 275 KKRILHC 281
RESULT 5
Q96QL4 PRELIMINARY; PRT; 296 AA.
ID Q96QL4
AC Q96QL4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:14706).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008456; AAH08456.1; -;
SQ SEQUENCE 296 AA; 33428 MW; 2CF2AA7FICDDB707 CRC64;
Query Match 92.7%; Score 38; DB 4; Length 296;
Best Local Similarity 85.7%; Pred. No. 2.2;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 KKRIMHC 7
    |||||
Db 275 KKRILHC 281

RESULT 6
Q9DD21
ID Q9DD21 PRELIMINARY; PRT; 303 AA.
AC Q9DD21
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 0610006A11RIK
GN 0610006A11RIK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavola H.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schrim L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Borls A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002231; BAB21950.1;
DR MGI; MGI:1915580; 0610006A11RIK.
DR InterPro; IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
SQ SEQUENCE 303 AA; 34491 MW; 52D8A3B02EE956EB CRC64;

Query Match 92.7%; Score 38; DB 11; Length 303;
Best Local Similarity 85.7%; Pred. No. 2.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KKRIMHC 7
    |||||
Db 274 KKRILHC 280

RESULT 7
Q9XZM5
ID Q9XZM5 PRELIMINARY; PRT; 106 AA.
AC Q9XZM5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RIBOSOMAL PROTEIN L44.
GN RPL44.
OS Leishmania amazonensis.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5659;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LV78;

Porter-Kelley J., Chaudhuri G.;
RT "Cloning and characterization of Leishmania ribosomal protein L44.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L44E FAMILY OF RIBOSOMAL PROTEINS.
DR EMBL; AF148853; AAD31928.2; -.
DR InterPro; IPR000552; Ribosomal_L44E.
DR Pfam; PF00935; Ribosomal_L44; 1.
DR ProDom; PD002841; Ribosomal_L44E; 1.
DR PROSITE; PS01172; RIBOSOMAL_L44E; 1.
KW Ribosomal Protein.
SQ SEQUENCE 106 AA; 12283 MW; F30A3AB2047B0334 CRC64;

Query Match 90.2%; Score 37; DB 5; Length 106;
Best Local Similarity 71.4%; Pred. No. 1.5;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KKRIMHC 7
    |||||
Db 6 KKRVMHC 12

RESULT 8
Q9LML0
ID Q9LML0 PRELIMINARY; PRT; 512 AA.
AC Q9LML0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F10K1.11 PROTEIN.
GN F10K1.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu S.X., Chan A., Yu G., Etgu P., Lee J.M., Lenz C., Pham P.,
RA Sakano H., Toriumi M., Chung M., Goldsmith A., Liu A., Liu A.,
RA Vaysberg M., Alcafi H., Brooks S., Suehrer E., Chao Q., Conn L.,
RA Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC F10K1 from Arabidopsis thaliana chromosome 1.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- COPACTOR: FAD (BY SIMILARITY).
DR EMBL; AC067971; AAF82202.1; -.
DR InterPro; IPR001327; FAD_pyr_redox.
DR Pfam; PF00070; pyr_redox; 1.
KW FAD; Flavoptrotein; Oxidoreductase.
SQ SEQUENCE 512 AA; 56857 MW; 1F63AFA9A1A2C13B CRC64;

Query Match 87.8%; Score 36; DB 10; Length 512;
Best Local Similarity 71.4%; Pred. No. 9.5;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KKRIMHC 7
    |||||
Db 236 KKRLLHC 242

RESULT 9
Q9D6S2
ID Q9D6S2 PRELIMINARY; PRT; 508 AA.
AC Q9D6S2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 2'-5' OLIGOADENYLATE SYNTHETASE-LIKE.
GN OASL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukuishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010034; BAB26655.1; -.
DR MGD; MGI:1344390; Oasi.
DR InterPro; IPR001797; 25A_synth.
DR InterPro; IPR001201; PAP_25A_core.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00213; UBQ; 1.
DR PROSITE; PS00833; 25A_SYNTH_2; UNKNOWN_1.
DR PROSITE; PS0153; 25A_SYNTH_3; 1.
DR PROSITE; PS00053; UBIQUITIN_2; 1.
SQ SEQUENCE 508 AA; 58767 MW; 9552B4540CC801A0 CRC64;

Query Match 85.4%; Score 35; DB 11; Length 508;
Best Local Similarity 71.4%; Pred. No. 15;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
Db 109 KKRILHC 115
|||||
|:|::||

RESULT 10
Q9QXV3 PRELIMINARY; PRT; 279 AA.
AC Q9QXV3; Q9QXV4; Q9QUP8; Q9QZX3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE INCI PROTEIN.
GN INCI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=129/SVJ; TISSUE=BRAIN, EMBRYONIC FIBROBLAST, AND SPLEEN;
RX MEDLINE=20011419; PubMed=10542254;
RA Zermanski M., Hill J.E., Kwek S.S., Grigorian I.A., Gurova K.V.,
RA Garkavtsev I.V., Diatchenko L., Koonin E.V., Gudkov A.V.;
RT "Structure and regulation of the mouse Ingi gene. Three alternative
RT transcripts encode two PHD finger proteins that have opposite effects
RT on p53 function.";
RL J. Biol. Chem. 274:32172-32181(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Rancourt D., Garkavtsev I.;
RT "Structural organization and expression pattern of the murine Ingi

```

```

RT gene.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE-BREAST TUMOR;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ISOFORM 1 INHIBITS P53-DEPENDENT TRANSCRIPTIONAL
CC ACTIVATION AND MAY FUNCTION AS AN ONCOPROTEIN. ISOFORM 2 ACTS AS A
CC NEGATIVE GROWTH REGULATOR BY COOPERATING WITH P53 IN
CC TRANSCRIPTIONAL ACTIVATION OF P53-RESPONSIVE GENES AND MAY ACT AS
CC A TUMOR SUPPRESSOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, WIDELY EXPRESSED WITH HIGHEST
CC LEVELS IN THYMUS AND TESTIS. EXPRESSED THROUGHOUT THE WHOLE EMBRYO
CC AT ALL STAGES OF DEVELOPMENT EXAMINED. AT DAY 10, HIGHEST
CC EXPRESSION IS FOUND IN THE YOLK SAC WHILE AT DAY 16 AND 18, HIGHER
CC LEVELS ARE FOUND IN INNER COMPARTMENTS OF BONE.
CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST EXPRESSION OF ISOFORM
CC 1 IS FOUND AT DAY 11 WHILE HIGHEST EXPRESSION OF ISOFORM 2 IS
CC FOUND AT DAY 7.
CC -1- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.
DR EMBL; AF177753; AAF16911.1; -.
DR EMBL; AF177755; AAF16908.1; -.
DR EMBL; AF177756; AAF16909.1; -.
DR EMBL; AF177757; AAF16910.1; -.
DR EMBL; AF149820; AAF09183.1; -.
DR EMBL; BC016573; AAH16573.1; -.
DR MGD; MGI:1349481; Ingi.
DR InterPro; IPR001965; PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
KW Oncogene; Anti-oncogene; Alternative splicing.
FT VARSPPLIC 210 259 PHD-FINGER.
FT CONFLICT 203 203 MISSING (IN ISOFORM 2).
FT CONFLICT 203 203 L -> F (IN REF. 2).
SQ SEQUENCE 279 AA; 32109 MW; 6765C984EEF179F4 CRC64;

Query Match 82.9%; Score 34; DB 11; Length 279;
Best Local Similarity 57.1%; Pred. No. 15;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
Db 67 KRRVLHC 73
|||||
|:|::||

RESULT 11
Q9M149 PRELIMINARY; PRT; 401 AA.
AC Q9M149;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE PHOSPHATIDYLINOSITOL KINASE.
GN AT4G01190.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161491; CAB80928.1; -.

```

DR InterPro: IPR002498; PIP5K.
DR Pfam: PF01504; PIP5K; 1.
DR SMART: SM00330; PIP5K; 1.
KW Kinase.
SQ SEQUENCE 401 AA; 45659 MW; 8A12D10DA2DED4CA CRC64;

Query Match 82.9%; Score 34; DB 10; Length 401;
Best Local Similarity 85.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
|||||

Db 353 KKRIMHC 359

RESULT 12

ID 004613 PRELIMINARY; PRT; 431 AA.

AC 004613;

DT 01-JUL-1997 (TREMBlrel. 04, Created)

DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE A_IG002N01.9 PROTEIN.

GN A_IG002N01.9.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

GN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Scheet P., Maggi L.;

RT "The sequence of A. thaliana IG002N01.";

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Wash-U;

RT "The A. thaliana Genome Sequencing Project.";

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Waterston R.;

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF007269; AAB61030.1; -

DR InterPro: IPR002498; PIP5K.

DR Pfam: PF01504; PIP5K; 1.

DR SMART: SM00330; PIP5K; 1.

SQ SEQUENCE 431 AA; 49356 MW; 07A53F23BDD942B CRC64;

Query Match

Best Local Similarity 82.9%; Score 34; DB 10; Length 431;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
|||||

Db 383 KKRIMHC 389

RESULT 13

Q03949

ID Q03949 PRELIMINARY; PRT; 189 AA.

AC Q03949;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE INSERTION ELEMENT IS895 HYPOTHETICAL 21.9 KDA PROTEIN (ORF1).

OS Anabaena sp. (strain PCC 7120).

OC Bacteria; Cyanobacteria; Nostocales; Nostoc.

OX NCBI_TaxID=103690;

RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=91358370; PubMed=1653219;
RA Alam J., Vrba J.M., Cai Y., Martin J.A., Weislo L.J., Curtis S.E.;
RT "Characterization of the IS895 family of insertion sequences from the
cyanobacterium Anabaena sp. strain PCC 7120.";
RL J. Bacteriol. 173:5778-5783(1991).
DR EMBL: M67475; AAA98138.1; -
DR PIR: A38117; A38117.
KW Hypothetical protein; Transposable element.
SQ SEQUENCE 189 AA; 21937 MW; B873A342856C2103 CRC64;

Query Match 80.5%; Score 33; DB 2; Length 189;

Best Local Similarity 71.4%; Pred. No. 17;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
|||||

Db 171 KKRIMHC 177

RESULT 14

Q9AKS0

ID Q9AKS0 PRELIMINARY; PRT; 267 AA.

AC Q9AKS0;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)

DE TATC PROTEIN.

GN TATC.

OS Pseudomonas stutzeri (Pseudomonas perfectomarina).

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas

OX NCBI_TaxID=316;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ZOBELL ATCC14405;

RX MEDLINE=21101859; PubMed=11160097;

RA Heikkilae M.P., Honisch U., Wunsch P., Zumft W.G.;

RT "Role of the Tat transport system in nitrous oxide reductase

translocation and cytochrome cdi biosynthesis in Pseudomonas

stutzeri.";

RL J. Bacteriol. 183:1663-1671(2001).

DR EMBL: AJ299712; CAC29149.1; -

DR InterPro: IPR002033; UPF0032.

DR Pfam: PF00902; UPF0032; 1.

DR PROSITE: PS01218; UPF0032; 1.

SQ SEQUENCE 267 AA; 29439 MW; 4369EB5E38BCD736 CRC64;

Query Match

Best Local Similarity 57.1%; Score 33; DB 2; Length 267;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
|||||

Db 21 KKRIMHC 27

RESULT 15

Q9ST63

ID Q9ST63 PRELIMINARY; PRT; 495 AA.

AC Q9ST63;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE PUTATIVE INTERNAL ROTENONE-INSENSITIVE NADH DEHYDROGENASE.

GN NDA1.

OS Solanum tuberosum (Potato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4113;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. DESIREE; TISSUE=LEAF;
RX MEDLINE=20040052; PubMed=10571867;
RA Rasmussen A.G., Svensson A., Knoop V., Grohmann L., Brennicke A.;
RT "Homologues of yeast and bacterial rotenone-insensitive NADH
dehydrogenases in higher eukaryotes: two enzymes are present in potato
mitochondria.";
RL Plant J. 20:79-87(1999).
CC -I- COFACTOR: FAD (BY SIMILARITY).
DR EMBL: AJ245861; CAB52796.1; -.
DR InterPro: IPR001327; FAD_pyr_redox.
DR Pfam: PF00070; pyr_redox; 1.
KW FAD; Flavoprotein; NAD; Oxidoreductase.
SQ SEQUENCE 495 AA; 54902 MW; 6AFC807BEB01340 CRC64;

Query Match 80.5%; Score 33; DB 10; Length 495;
Best Local Similarity 57.1%; Pred. No. 40;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|:|::|
Db 224 KRLLHC 230

Search completed: November 1, 2002, 14:49:46
Job time : 3.2717 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:41:32 : Search time 0.360129 seconds
(without alignments)
752.611 Million cell updates/sec

Title: US-09-606-129A-18
Perfect score: 41
Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	41	100.0	295	1 BIEA_RAT	P46844 rattus norv
2	38	92.7	296	1 BIEA_HUMAN	P53004 homo sapien
3	35	85.4	473	1 OASL_MOUSE	Q922f2 mus musculus
4	33	80.5	925	1 PIP1_YEAST	P40020 saccharomyc
5	32	78.0	140	1 FKB2_MOUSE	P45878 mus musculus
6	32	78.0	141	1 FKB2_HUMAN	P26885 homo sapien
7	32	78.0	328	1 HYBA_ECOLI	P37179 escherichia
8	32	78.0	528	1 DYR2_HUMAN	Q92630 homo sapien
9	32	78.0	571	1 AMPL_LYCES	Q10712 lycopersico
10	32	78.0	573	1 AMPL_SOLTU	P31427 solanum tub
11	31	75.6	105	1 RL44_TRYBB	P17843 trypanosoma
12	31	75.6	279	1 Y091_NPVOP	Q10341 oryza pseu
13	31	75.6	323	1 CYCH_XENLA	P51947 xenopus lae
14	31	75.6	451	1 SUN_HAEIN	P44788 haemophilus
15	31	75.6	768	1 CUL3_HUMAN	Q13618 homo sapien
16	31	75.6	768	1 CUL3_MOUSE	Q9jlv5 mus musculus
17	31	75.6	861	1 UL52_HSVJ7	P52468 human herpe
18	30	73.2	108	1 YZ09_MERJA	Q50271 methanococc
19	30	73.2	4128	1 PRKO_MOUSE	P97313 mus musculus
20	29	70.7	38	1 RL36_THEMA	Q9x116 thermotoga
21	29	70.7	142	1 SEC8_BUCAI	P57161 buchnera ap
22	29	70.7	144	1 YHP5_YEAST	P38808 saccharomyc
23	29	70.7	257	1 ZG1L_CAEEL	P55326 caenorhabdi
24	29	70.7	300	1 RANT_BPF22	P03037 bacterioph
25	29	70.7	359	1 ODPB_RAT	P49432 rattus norv
26	29	70.7	423	1 SHP1_YEAST	P34223 saccharomyc
27	29	70.7	560	1 DY2H_SCHPO	Q10264 schizosacch
28	29	70.7	588	1 DYR3_HUMAN	Q43781 homo sapien
29	29	70.7	632	1 AFUB_HAEIN	Q57341 haemophilus
30	29	70.7	670	1 REP_HAEIN	P44804 haemophilus
31	29	70.7	918	1 YK62_CAEEL	P34341 caenorhabdi
32	29	70.7	1059	1 CAPU_DROME	Q24120 drosophila
33	29	70.7	1790	1 SEPA_EMENI	P78621 emericecella

34 28 68.3 113 1 FLIT_BACSU P39740 bacillus su
35 28 68.3 172 1 IPYR_RICPR Q9zcw5 rickettsia
36 28 68.3 249 1 YAT6_RHOBL P05449 rhodospaudo
37 28 68.3 313 1 YOAM_BACSU P45910 bacillus su
38 28 68.3 324 1 IDSA_METTM Q53479 methanobact
39 28 68.3 335 1 UL16_HSV6U P24442 human herpe
40 28 68.3 356 1 GBAL_DICTDI P16894 dictyosteli
41 28 68.3 428 1 DGTP_MYCSM O52199 mycobacteri
42 28 68.3 459 1 NEK3_HUMAN P51956 homo sapien
43 28 68.3 477 1 URT1_DESRO P98119 desmodus ro
44 28 68.3 477 1 URT2_DESRO P15638 desmodus ro
45 28 68.3 494 1 GPPA_ECOLI P25552 escherichia

ALIGNMENTS

RESULT 1
BIEA_RAT
ID BIEA_RAT STANDARD; PRT; 295 AA.
AC P46844;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Biliverdin reductase A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-
reductase).
DE BLVRA OR BLVR.
GN Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=92156147; PubMed=1371282;
RA Fakhrai H., Maines M.D.;
RT "Expression and characterization of a cDNA for rat kidney biliverdin
reductase. Evidence suggesting the liver and kidney enzymes are the
same transcript product".
RT J. Biol. Chem. 267:4023-4029(1992).
RN [2]
RP MUTAGENESIS.
RX MEDLINE=94291657; PubMed=8020496;
RA McCoubrey W.K. Jr., Maines M.D.;
RT "Site-directed mutagenesis of cysteine residues in biliverdin
reductase. Roles in substrate and cofactor binding".
RL Eur. J. Biochem. 222:597-603(1994).
CC -|- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN; DISPLAYS TWO DISTINCT
PH OPTIMA USING A DIFFERENT COFACTOR AT EACH PH: NADH AT THE LOWER
PH 6.7-6.9 RANGE AND NADPH AT PH 8.5-8.7. NADPH, HOWEVER, IS THE
PROBABLE COFACTOR IN BIOLOGICAL SYSTEMS.
CC -|- CATALYTIC ACTIVITY: Bilirubin + NAD(P)(+) = biliverdin + NAD(P)H.
CC -|- COFACTOR: BINDS ONE ZINC ION.
CC -|- PATHWAY: FINAL STEP IN HEME METABOLISM.
CC -|- SUBUNIT: MONOMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: TO E.COLI YHHX.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/announcements/](http://www.isb-sib.ch/announcements)
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81681; AAA40830.1; -
CC InterPro: IPR000683; GFO_IDH_MOCA.
DR Pfam; PF01408; GFO_IDH_MOCA; 1.
KW Oxidoreductase; NAD; NADP; Zinc.
FT PROPEP 1 2
FT CHAIN 3 295 BILIVERDIN REDUCTASE A.
FT DOMAIN 11 16 POLY-VAL.

```

FT METAL 279 279 ZINC (POTENTIAL).
FT METAL 280 280 ZINC (POTENTIAL).
FT METAL 291 291 ZINC (POTENTIAL).
FT METAL 292 292 ZINC (POTENTIAL).
FT MUTAGEN 73 73 C->A: LOSS OF ACTIVITY.
FT MUTAGEN 280 280 C->A: REDUCED ACTIVITY.
FT MUTAGEN 291 291 C->A: REDUCED ACTIVITY.
SQ SEQUENCE 295 AA; 33565 MW; 219C8EA96C150588 CRC64;

Query Match 100.0%; Score 41; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
    |||||
Db 274 KKRIMHC 280

RESULT 2
BIEA_HUMAN
ID BIEA_HUMAN STANDARD; PRT; 296 AA.
AC P53004;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE BILIVERDIN REDUCTASE A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-
DE reductase).
GN BLVRA OR BLVR OR BVR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96202961; PubMed=8631357;
RA Maines M.D., Polevoda B.V., Huang T.-J., McCoubrey W.K. Jr.;
RT "Human biliverdin IXalpha reductase is a zinc-metalloprotein.
RT Characterization of purified and Escherichia coli expressed
RT enzymes";
RL Eur. J. Biochem. 235:372-381(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Komuro A., Tobe T., Nakano Y., Yamaguchi T., Tomita M.;
RP Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-117 FROM N.A.
RA Cordes M., Wollam C., Carter T.;
RP Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 3-36; 48-74 AND 228-248.
RC TISSUE=Liver;
RX MEDLINE=93143333; PubMed=8424666;
RA Maines M.D., Trakshel G.M.;
RT "Purification and characterization of human biliverdin reductase.";
RL Arch. Biochem. Biophys. 300:320-326(1993).
RN [5]
RP SEQUENCE OF 3-22.
RC TISSUE=Liver;
RX MEDLINE=95014177; PubMed=7929092;
RA Yamaguchi T., Komoda Y., Nakajima H.;
RT "Biliverdin-IX alpha reductase and biliverdin-IX beta reductase from
RT human liver. Purification and characterization.";
RL J. Biol. Chem. 269:24343-24348(1994).
CC -1- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN.
CC -1- CATALYTIC ACTIVITY: Bilirubin + NAD(P)(+) = biliverdin + NAD(P)H.
CC -1- COFACTOR: BINDS ONE ZINC ION. HAS DUAL PH/COFACTOR (NADH, NADPH)
CC SPECIFICITY. USES NADH AT THE ACIDIC PH RANGE (6-6.7) AND NADPH AT
CC THE ALKALINE RANGE (8.5-8.7).
CC -1- PATHWAY: FINAL STEP IN HEME METABOLISM.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: LIVER.

```

```

CC -1- SIMILARITY: TO E.COLI YHHX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X93086; CAA63635.1; -
DR EMBL; U34877; AAC35588.1; -
DR EMBL; AC005189; AAC25526.1; -
DR MIM; 109750; -
DR InterPro; IPR000683; GFO_IDH_MoCA.
DR Pfam; PF01408; GFO_IDH_MoCA; 1.
KW Oxidoreductase; NAD; NADP; Zinc.
FT PROPEP 1 2
FT CHAIN 3 296 BILIVERDIN REDUCTASE A.
FT DOMAIN 11 16 POLY-VAL.
FT METAL 280 280 ZINC (POTENTIAL).
FT METAL 281 281 ZINC (POTENTIAL).
FT METAL 292 292 ZINC (POTENTIAL).
FT METAL 293 293 ZINC (POTENTIAL).
FT CONFLICT 3 3 A -> T (IN REF. 2).
FT CONFLICT 154 155 SD -> AG (IN REF. 2).
FT CONFLICT 160 160 D -> E (IN REF. 2).
SQ SEQUENCE 296 AA; 33488 MW; 0DFD3B386FADF0A CRC64;

Query Match 92.7%; Score 38; DB 1; Length 296;
Best Local Similarity 85.7%; Pred. No. 0.62;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
    |||||
Db 275 KKRILHC 281

RESULT 3
OASL_MOUSE
ID OASL_MOUSE STANDARD; PRT; 473 AA.
AC Q922F2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 54 kDa 2'-5'-oligoadenylate synthetase like protein (EC 2.7.7.-) (p54
DE OASL) (P54OASL) (M1204).
GN OASL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=99323969; PubMed=10395668;
RA Tiefenthaler M., Marksteiner R., Neyer S., Koch F., Hofer S.,
RA Schuler G., Nussenzweig M., Schneider R., Heufner C.;
RT "M1204, a novel 2',5' oligoadenylate synthetase with a ubiquitin-like
RT extension, is induced during maturation of murine dendritic cells.";
RL J. Immunol. 163:760-765(1999).
CC -1- FUNCTION: MAY PLAY A ROLE IN MEDIATING RESISTANCE TO VIRUS
CC INFECTION, CONTROL OF CELL GROWTH, DIFFERENTIATION, AND APOPTOSIS.
CC -1- CATALYTIC ACTIVITY: BINDS DOUBLE-STRANDED RNA AND POLYMERIZES ATP
CC INTO PP(A2'P5'A)N OLIGOMERS, WHICH ACTIVATE THE LATENT RNASE L
CC THAT, WHEN ACTIVATED, CLEAVES SINGLE-STRANDED RNAS.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SPLEEN DENDRITIC CELLS,
CC WHEREAS, IN BONE MARROW-DERIVED DENDRITIC CELLS, THE AMOUNT
CC INCREASES DURING THE MATURATION PROCESS. EXPRESSED IN MANY ORGANS,
CC THE HIGHEST LEVELS BEING IN THYMUS, LUNG, AND BONE MARROW.
CC -1- SIMILARITY: BELONGS TO THE 2-5A SYNTHETASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 UBIQUITIN-LIKE DOMAIN.
CC -1- CAUTION: THIS MAY NOT BE THE TRUE ORTHOLOG OF HUMAN OASL.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: AF068835; AAD02818.1; -;
CC MGI: 1344390; Qas1.
CC InterPro: IPR001797; 25A_Synth.
CC InterPro: IPR001201; PAP_25A_core.
CC InterPro: IPR000626; Ubiquitin.
CC Pfam: PF00240; ubiquitin; 1.
CC PROSITE: PS00832; 25A_SYNTH_1; FALSE_NEG.
CC PROSITE: PS00833; 25A_SYNTH_2; 1.
CC PROSITE: PS0152; 25A_SYNTH_3; 1.
CC PROSITE: PS0053; UBIQUITIN_2; FALSE_NEG.
CC RNA-binding; Transferase; Nucleotidyltransferase.
CC DOMAIN 435 473 UBIQUITIN-LIKE.
CC SEQUENCE 473 AA; 54625 MW; 570E0E08A51C8460 CRC64;

Query Match 85.4%; Score 35; DB 1; Length 473;
Best Local Similarity 71.4%; Pred. No. 4.2;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
Db 109 KKRLIHC 115

RESULT 4
PIPL_YEAST
ID PIPL_YEAST STANDARD; PRT; 925 AA.
AC P40020;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Polymerase-interacting protein 1 (Factor interacting with REF).
GN PIPL OR FIR1 OR YER032W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W303;
RA del Olmo M., Gross S., Moore C.L.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S., Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D., Moresdale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C., Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V., Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE-97339480; PubMed-9196079;
RA Ruskak R., Pereira S., Platt T.;
RT "RNA binding analysis of yeast REF2 and its two-hybrid interaction with a new gene product, FIR1";
RL Gene Expr. 6:241-258(1996).
CC -!- FUNCTION: Interacts with poly(A) polymerase and with REF2.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: U17262; AAB46625.1; -;
CC EMBL: U18778; AAB64565.1; -;
CC SGD: S0000834; FIR1.
CC CONFLICT 663 663 R -> P (IN REF. 2).
CC SEQUENCE 925 AA; 104701 MW; 707D9839EE31322B CRC64;

Query Match 80.5%; Score 33; DB 1; Length 925;
Best Local Similarity 71.4%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
Db 841 KKRLSHC 847

RESULT 5
FKB2_MOUSE
ID FKB2_MOUSE STANDARD; PRT; 140 AA.
AC P45878;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE FK506-binding protein precursor (FKBP-13) (peptidyl-prolyl cis-trans isomerase) (FKBP) (EC 5.2.1.8).
GN FKBP2 OR FKBP13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ; TISSUE=Liver;
RX MEDLINE=94085790; PubMed=7505249;
RA Hendrickson B.A., Zhang W., Craig R.J., Jin Y.J., Bierer R.E., Burakoff S.J., Dilella A.G.;
RT "Structural organization of the genes encoding human and murine FK506-binding protein (FKBP) 13 and comparison to FKBP1.";
RL Gene 134:271-275(1993).
CC -!- FUNCTION: PPIases accelerate the folding of proteins. May function as a component of membrane cytoskeletal scaffolds.
CC -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -!- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC -!- SUBUNIT: Interacts with the C-terminal domain of 4.1G.
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN. MEMBRANE ASSOCIATED (PROBABLY).
CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: M77831; AAA37631.1; -;
CC HSP; P20081; 1YAT.
CC MGI: 95542; FKBP2.
CC InterPro: IPR001179; FKBP_PPIase.
CC Pfam: PF00254; FKBP; 1.
CC PROSITE: PS00453; FKBP_PPIASE_1; 1.
CC PROSITE: PS00454; FKBP_PPIASE_2; 1.
CC PROSITE: PS50059; FKBP_PPIASE_3; 1.
CC Isomerase; Rotamase; Signal; Endoplasmic reticulum.
CC SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 140 FK506-BINDING PROTEIN.
CC SITE 137 140 PREVENT SECRETION FROM ER (POTENTIAL).

```
SQ SEQUENCE 140 AA; 15344 MW; F4E7FCC7766A0416 CRC64;
Query Match 78.0%; Score 32; DB 1; Length 140;
Best Local Similarity 71.4%; Pred. No. 5.4;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKRIMHC 7
DB 34 KKRVDHC 40

RESULT 6
FKB2_HUMAN
ID FKB2_HUMAN STANDARD; PRT; 141 AA.
AC P26885;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE FK506-binding protein precursor (FKBP-13) (peptidyl-prolyl cis-trans
DE isomerase) (PPIase) (EC 5.2.1.8).
GN FKBP2 OR FKBP13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Colon carcinoma;
RX MEDLINE=91319747; PubMed=1713687;
RA Jin Y.-J., Albers M.W., Lane W.S., Blier B.E., Schreiber S.L.,
RA Burakoff S.J.;
RT "Molecular cloning of a membrane-associated human FK506- and
RT rapamycin-binding protein, FKBP-13."
BL proc. Natl. Acad. Sci. U.S.A. 88:6677-6681(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93112052; PubMed=1281998;
RA Dilella A.G., Hawkins A., Craig R.J., Schreiber S.L., Griffin C.A.;
RT "Chromosomal band assignments of the genes encoding human FKBP12 and
RT FKBP13."
RL Biochem. Biophys. Res. Commun. 189:819-823(1992).
CC -1- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN. MEMBRANE
CC ASSOCIATED (PROBABLE).
CC -1- TISSUE SPECIFICITY: T-CELLS AND THYMUS.
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M65128; AAA58473.1; -
CC EMBL; M75099; AAA36563.1; -
CC PIR; JC1365; JC1365.
CC HSSP; Q00688; 1PBK.
CC MIN; 186946; -
CC InterPro; IPR001179; FKBP_PPIase.
CC Pfam; PF00254; FKBP; 1.
CC PROSITE; PS00453; FKBP_PPIASE_1; 1.
CC PROSITE; PS00454; FKBP_PPIASE_2; 1.
CC PROSITE; PS00509; FKBP_PPIASE_3; 1.
KW Isomerase; Rotamase; Signal; Endoplasmic reticulum; Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 141 FK506-BINDING PROTEIN.
FT SITE 138 141 PREVENT SECRETION FROM ER (POTENTIAL).
FT VARIANT 21 21 S -> TA.
```

```
FT VARIANT 24 24 /FTid=VAR_006410.
FT T -> A.
FT /FTid=VAR_006411.
FT Y -> C.
FT /FTid=VAR_006412.
SQ SEQUENCE 141 AA; 15654 MW; 9F4751CA7D82D064 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 141;
Best Local Similarity 71.4%; Pred. No. 5.4;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKRIMHC 7
DB 35 KKRVDHC 41

RESULT 7
HYBA_ECOLI
ID HYBA_ECOLI STANDARD; PRT; 328 AA.
AC P37179;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE Hydrogenase-2 operon protein hyba precursor.
DE HYBA OR B2996 OR Z4350 OR ECS3881.
GN Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / TGI;
RX MEDLINE=94292472; PubMed=8021226;
RA Menon N.K., Chatelus C.Y., Dervartanian M., Wendt J.C.,
RA Shanmugam K.T., Peck H.D. Jr., Przybyla A.E.;
RT "Cloning, sequencing, and mutational analysis of the hyb operon
RT encoding Escherichia coli hydrogenase 2."
RL J. Bacteriol. 176:4416-4423(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: PARTICIPATES IN THE PERIPLASMIC ELECTRON-TRANSFERRING
```

```
CC ACTIVITY OF HYDROGENASE 2 DURING ITS CATALYTIC TURNOVER.
CC -1- COFACTOR: BINDS 3 4FE-4S CLUSTERS AND A 3FE-4S CLUSTER.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE THE SMALL SUBUNIT
CC OF HYDROGENASE 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U09177; AAA21589.1; -.
CC EMBL; U28377; AAA69163.1; -.
CC EMBL; AE000382; AAC76032.1; -.
CC EMBL; AE005529; AAG58133.1; -.
CC EMBL; AP002363; BAB37304.1; -.
CC HSSP; P55907; 1XER.
CC Ecogene; EGI1799; Hyba.
CC InterPro; IPR001450; 4Fe4S_ferredoxin.
CC Pfam; PF00037; fer4; 1.
CC PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
CC Oxidoreductase; Signal; Periplasmic; Iron-sulfur; 4Fe-4S; 3Fe-4S;
CC Complete proteome.
CC -----
FT CHAIN 1 27 POTENTIAL.
FT METAL 28 328 HYDROGENASE-2 OPERON PROTEIN HYBA.
FT METAL 47 47 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 50 50 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 53 53 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 57 57 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 112 112 IRON-SULFUR 2 (3FE-4S) (POTENTIAL).
FT METAL 115 115 IRON-SULFUR 2 (3FE-4S) (POTENTIAL).
FT METAL 120 120 IRON-SULFUR 2 (3FE-4S) (POTENTIAL).
FT METAL 124 124 IRON-SULFUR 2 (3FE-4S) (POTENTIAL).
FT METAL 145 145 IRON-SULFUR 3 (4FE-4S) (POTENTIAL).
FT METAL 148 148 IRON-SULFUR 3 (4FE-4S) (POTENTIAL).
FT METAL 151 151 IRON-SULFUR 3 (4FE-4S) (POTENTIAL).
FT METAL 155 155 IRON-SULFUR 3 (4FE-4S) (POTENTIAL).
FT METAL 174 174 IRON-SULFUR 4 (4FE-4S) (POTENTIAL).
FT METAL 177 177 IRON-SULFUR 4 (4FE-4S) (POTENTIAL).
FT METAL 193 193 IRON-SULFUR 4 (4FE-4S) (POTENTIAL).
FT METAL 197 197 IRON-SULFUR 4 (4FE-4S) (POTENTIAL).
CC SEQUENCE 328 AA; 36003 MW; 77203A0F50F61662 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 328;
Best Local Similarity 71.4%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
Db 109 KKQCMHC 115

RESULT 8
DTR2_HUMAN
ID DTR2_HUMAN STANDARD; PRT; 528 AA.
AC Q92630;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 2
DE (EC 2.7.1.-).
GN DYRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=Brain;
RX MEDLINE=98421512; PubMed=9748265;
RA Becker W., Weber Y., Wetzel K., Eirmbter K., Tejedor F.J.,
RA Joost H.-G.;
RT "Sequence characteristics, subcellular localization, and substrate
RT specificity of DYRK-related kinases, a novel family of dual
RL J. Biol. Chem. 273:25893-25902(1998).
RN [2]
RP SEQUENCE OF 320-528 FROM N.A.
RC TISSUE=Placenta;
RA Becker W., Joost H.-G.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN VITRO; CAN PHOSPHORYLATE HISTONES H3 AND H2B ON SER
CC AND THR RESIDUES. MAY BE INVOLVED IN THE REGULATION OF CELLULAR
CC GROWTH AND/OR DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MNB/DYRK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y13493; CAA73885.1; -.
CC EMBL; Y09216; CAA70418.1; -.
CC HSSP; Q16539; 1WFC.
CC MIM; 603496; -.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002290; Ser_thr_pkinase.
CC Pfam; PF00069; pkinase; 2.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
CC ATP-binding; Phosphorylation.
CC DOMAIN 149 462 PROTEIN KINASE.
CC NP_BIND 155 163 ATP (BY SIMILARITY).
CC BINDING 178 178 ATP (BY SIMILARITY).
CC ACT_SITE 275 275 BY SIMILARITY.
CC SEQUENCE 528 AA; 59714 MW; AF2C6822ED952D7 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 528;
Best Local Similarity 71.4%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
Db 268 KNRIHC 274

RESULT 9
AMPI_LYCES
ID AMPI_LYCES STANDARD; PRT; 571 AA.
AC Q10712; Q9S9A3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Aminopeptidase 1, chloroplast precursor (EC 3.4.11.1) (Leucine
DE aminopeptidase) (LAP) (Leucyl aminopeptidase) (Proline aminopeptidase)
DE (EC 3.4.11.5) (Poly1 aminopeptidase) (DR57).
GN LAP1 OR LAP OR LAP2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RP SEQUENCE FROM N.A.
```

RN SEQUENCE FROM N.A.
 RP STRAIN=CV. PETO 238R; TISSUE=Leaf;
 RX MEDLINE=96421572; PubMed=8824220;
 RA Gu Y.Q., Chao W.S., Walling L.L.;
 RT "Localization and post-translational processing of the wound-induced
 RL leucine aminopeptidase proteins of tomato.";
 RL J. Biol. Chem. 271:25880-25887(1996).
 RN [2]
 RP SEQUENCE OF 49-571 FROM N.A.
 RC STRAIN=CV. VF36; TISSUE=Platyl;
 RX MEDLINE=95375233; PubMed=7647301;
 RA Milligan S.B., Gasser C.S.;
 RT "Nature and regulation of pistil-expressed genes in tomato.";
 RL plant Mol. Biol. 28:691-711(1995).
 RN [3]
 RP SEQUENCE OF 103-571 FROM N.A.
 RC STRAIN=CV. PETO 238R; TISSUE=Leaf;
 RX MEDLINE=94052201; PubMed=8234334;
 RA Pautot V., Holzer F.M., Reisch B., Walling L.L.;
 RT "Leucine aminopeptidase: an inducible component of the defense
 RL response in Lycopodium esculentum (tomato).";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:9906-9910(1993).
 CC -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
 CC TURNOVER OF INTRACELLULAR PROTEINS.
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
 CC including Pro although not Arg or Lys, and Xbb may be Pro.
 CC -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Chloroplast.
 CC -!- INDUCTION: BY WOUNDING.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17; ALSO KNOWN AS THE
 CC CYTOSOL AMINOPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U50151; AAC49456.1; -;
 DR EMBL; U50152; AAC49457.1; -;
 DR EMBL; U20593; AAA80498.1; -;
 DR HSSP; P00727; 1LAM.
 DR MEROPS; M17.002.
 DR InterPro; IPR000819; Peptidase_M17.
 DR Pfam; PF00883; Peptidase_M17; 1.
 DR Pfam; PF02789; Peptidase_M17_N; 1.
 DR PRINTS; PR00481; LAMNOPPTDASE.
 DR PROSITE; PS00631; CYTOSOLAP; 1.
 KW Transit peptide; Chloroplast; Aminopeptidase; Hydrolase; Zinc.
 FT TRANSIT 1 53 CHLOROPLAST (POTENTIAL).
 FT CHAIN 54 571 AMINOPEPTIDASE 1.
 FT DOMAIN 169 174 POLY-ALA.
 FT METAL 342 342 ZINC 2 (BY SIMILARITY).
 FT METAL 347 347 ZINC 1 AND 2 (BY SIMILARITY).
 FT METAL 367 367 ZINC 2 (BY SIMILARITY).
 FT METAL 427 427 ZINC 1 (BY SIMILARITY).
 FT METAL 429 429 ZINC 1 AND 2 (BY SIMILARITY).
 FT ACT_SITE 354 354 POTENTIAL.
 FT ACT_SITE 431 431 POTENTIAL.
 FT VARIANT 358 358 R -> G.
 FT CONFLICT 271 271 P -> N (IN REF. 2).
 FT CONFLICT 315 315 T -> S (IN REF. 3).
 FT CONFLICT 515 515 T -> L (IN CLONE PBLAP2).
 SQ SEQUENCE 571 AA; 60279 MW; C7A224837E73939D CRC64;

Query Match

78.08; Score 32; DB 1; Length 571;

 Best Local Similarity 83.38; Pred. No. 22; Indels 0; Gaps 0;
 Matches 5; Conservative 1; Mismatches 0;

QY 2 KRIMHC 7
 DB 48 KRIVHC 53
 RESULT 10
 AMPL_SOLUTU
 ID AMPL_SOLUTU STANDARD; PRT; 573 AA.
 AC P31427; 1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 41, Last annotation update)
 DE Aminopeptidase, chloroplast precursor (EC 3.4.11.1) (Leucine
 DE aminopeptidase) (LAP) (Leucyl aminopeptidase) (Proline aminopeptidase)
 DE (EC 3.4.11.5) (Prolyl aminopeptidase).
 GN LAP.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. DESIREE;
 RX MEDLINE=94339796; PubMed=7765119;
 RA Herbers K., Prat S., Willmitzer L.;
 RT "Functional analysis of a leucine aminopeptidase from Solanum
 RL tuberosum L.";
 RL Planta 194:230-240(1994).
 RN [2]
 RP SEQUENCE OF 19-573 FROM N.A.
 RC STRAIN=CV. DESIREE; TISSUE=Leaf;
 RX MEDLINE=93005746; PubMed=1392612;
 RA Hildmann T., Ebner M., Pena-Cortes H., Sanchez-Serrano J.J.,
 RA Willmitzer L., Prat S.;
 RT "General roles of abscisic and jasmonic acids in gene activation as a
 RT result of mechanical wounding.";
 RL Plant Cell 4:1157-1170(1992).
 CC -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
 CC TURNOVER OF INTRACELLULAR PROTEINS.
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
 CC including Pro although not Arg or Lys, and Xbb may be Pro.
 CC -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
 CC -!- SUBUNIT: HOMOHETEROMER (PROBABLE).
 CC -!- SUBCELLULAR LOCATION: Chloroplast (By similarity).
 CC -!- TISSUE SPECIFICITY: IN TUBERS AND FLORAL BUDS OF UNTREATED PLANTS.
 CC AFTER ABA TREATMENT OR MECHANICAL WOUNDING IS MOSTLY ACCUMULATED
 CC IN LEAVES, TO A LESSER EXTENT IN STEMS, BUT NOT IN ROOTS.
 CC -!- INDUCTION: BY ABSICISIC ACID (ABA), JASMONIC ACID (JA) AND
 CC WOUNDING.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17; ALSO KNOWN AS THE
 CC CYTOSOL AMINOPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X77015; CAA5A314.1; -;
 DR EMBL; X67845; CAA48038.1; -;
 DR PIR; S24769; S24769.
 DR PIR; PQ0470; PQ0470.
 DR HSSP; P00727; 1LAM.
 DR MEROPS; M17.002.
 DR InterPro; IPR000819; Peptidase_M17.
 DR Pfam; PF00883; Peptidase_M17; 1.
 DR Pfam; PF02789; Peptidase_M17_N; 1.
 DR PRINTS; PR00481; LAMNOPPTDASE.
 DR PROSITE; PS00631; CYTOSOLAP; 1.

KW Transit peptide; Chloroplast; Aminopeptidase; Hydrolase; Zinc.
FT CHAIN 1 53 CHLOROPLAST (POTENTIAL).
FT DOMAIN 169 174 AMINOPEPTIDASE.
FT METAL 342 347 POLY-ALA.
FT METAL 347 347 ZINC 2 (BY SIMILARITY).
FT METAL 367 367 ZINC 1 AND 2 (BY SIMILARITY).
FT METAL 427 427 ZINC 2 (BY SIMILARITY).
FT METAL 427 427 ZINC 1 (BY SIMILARITY).
FT METAL 429 429 ZINC 1 AND 2 (BY SIMILARITY).
FT ACT_SITE 354 354 POTENTIAL.
FT ACT_SITE 431 431 POTENTIAL.
SQ SEQUENCE 573 AA; 60122 MW; 31521454A47FB291 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 573;

Best Local Similarity 83.3%; Pred. No. 22;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRIMHC 7

DB 48 KRIVHC 53

RESULT 11

RL44_TRYBB STANDARD; PRT; 105 AA.
AC PL7843;
DT 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60S ribosomal protein L44.
GN RPL44.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EATRO 1125;
RX MEDLINE=90251460; PubMed=2339065;
RA Tebabi P., Halleux S., Pays E.;
RT "Nucleotide sequence of a full-length cDNA coding for the ribosomal L44 protein of Trypanosoma brucei."
RL Nucleic Acids Res. 18:2809-2809(1990).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE L44E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; X52122; CAA36367.1; -.
DR PIR; S10012; R6UT6A.
DR InterPro; IPR000552; Ribosomal_L44E.
DR Pfam; PF00935; Ribosomal_L44; 1.
DR ProDom; PD002841; Ribosomal_L44E; 1.
DR ProSite; PS01172; RIBOSOMAL_L44E; 1.
KW Ribosomal protein.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 105 AA; 12322 MW; FA19423F109E7819 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 105;

Best Local Similarity 71.4%; Pred. No. 6.5;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KRIMHC 7

DB 5 KKKMHC 11

RESULT 12

Y091_NPVOP STANDARD; PRT; 279 AA.
AC O10341;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 29.3 kDa protein (ORF92).
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; U75930; AAC59091.1; -.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 279 AA; 29289 MW; 6FA4DAA01009DBF0 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 279;

Best Local Similarity 66.7%; Pred. No. 17;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRIMHC 7

DB 233 QRVNHC 238

RESULT 13

CYCH_XENLA STANDARD; PRT; 323 AA.
AC P51947;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cyclin H (MO15-associated protein) (p36).
GN CCNH.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez A.-M.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PARTIAL SEQUENCE.
RC TISSUE=Oocyte;
RX MEDLINE=95045408; PubMed=7957080;
RA Labbe J.-C., Martnez A.-M., Fesquet D., Capony J.-P., Darbon J.-M.,
RA Derancourt J., Devault A., Morin N., Cavadore J.-C., Doree M.;
RT "P40MO15 associates with a p36 subunit and requires both nuclear
RT translocation and Thr176 phosphorylation to generate cdk-activating
RT kinase activity in Xenopus oocytes.";
RL EMBO J. 13:5153-5164(1994).
CC -1- FUNCTION: Regulates CDK7, the catalytic subunit of the CDK-
CC activating kinase (CAK) enzymatic complex. CAK activates the
CC cyclin-associated kinases CDC2/CDK1, CDK2, CDK4 and CDK6 by

CC threonine phosphorylation. CAK complexed to the core-TFIIF basal
 CC transcription factor activates RNA polymerase II by serine
 CC phosphorylation of the repetitive carboxyl-terminus domain (CTD)
 CC of its large subunit (POLR2A), allowing its escape from the
 CC promoter and elongation of the transcripts. Involved in cell cycle
 CC control and in RNA transcription by RNA polymerase II. Its
 CC expression and activity are constant throughout the cell cycle (By
 CC similarity).
 CC -!- SUBUNIT: Associates primarily with CDK7 and MAT1 to form the
 CC CAK complex. CAK can further associate with the core-TFIIF to
 CC form the TFIIF basal transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U20505; AAA62236.1; --
 DR HSP; P51946; IJWK.
 DR InterPro: IPR000553; Cyclin.
 DR Pfam: PF00134; cyclin; 1.
 DR SMART: SM00385; CYCLIN; 1.
 DR PROSITE: PS00292; CYCLINS; FALSE_NEG.
 KW Transcription regulation; Cell cycle; Nuclear protein; Cyclin;
 KW Phosphorylation.
 FT CONFLICT 266 266 R -> Y (IN REF. 2; AA SEQUENCE).
 SQ SEQUENCE 323 AA; 37600 MW; 14BCDCA00843DC8 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 323;
 Best Local Similarity 71.4%; Pred. No. 20;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
 III: II
 DB 274 KKRILHC 280

RESULT 14
 SUN_HAEIN STANDARD; PRT; 451 AA.
 AC P44789;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE SUN protein (FMU protein).
 GN SUN OR FMU OR HI0624.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RC SEQUENCE FROM N.A.
 PC STRAIN-RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Karvay A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Frichman J.L., Fuhrmann J.I., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -!- SIMILARITY: BELONGS TO THE SUN (BACTERIAL) / NUCLEOLAR PROTEIN
 CC NOLI/NOP2 (EUKARYOTES) FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U32745; AAC22284.1; --
 DR TIGR; HI0624; --
 DR InterPro: IPR001678; Noli_Nop2_Sun.
 DR InterPro: IPR000139; Nusb.
 DR InterPro: IPR000051; SAM_bind.
 DR Pfam: PF01189; Noli_Nop2_Sun; 1.
 DR Pfam: PF01029; Nusb; 1.
 DR PRODOM: PD005242; Nusb; 1.
 DR PROSITE: PS01153; NOLI_NOP2_SUN; 1.
 DR Complete proteome.
 KW SEQUENCE 451 AA; 50597 MW; D91FAB88FFDE34B0 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 451;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7
 III: II
 DB 88 KTRIVHC 94

RESULT 15
 CUL3_HUMAN STANDARD; PRT; 768 AA.
 AC Q13618; O75415; Q9UB18; Q9UET7;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cullin homolog 3 (CUL-3).
 GN CUL3 OR KIRA0617.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 PC MEDLINE=98406061; PubMed=9733711;
 Du M., Sansores-Garcia L., Zu Z., Wu K.K.;
 RT "Cloning and expression analysis of a novel salicylate suppressible
 RT gene, Hs-CUL-3, a member of cullin/Cdc53 family.";
 RL J. Biol. Chem. 273:24289-24292(1998).
 RN [2]
 RC SEQUENCE FROM N.A.
 PC TISSUE=Brain;
 RX MEDLINE=98403880; PubMed=9734811;
 Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:169-176(1998).
 RN [3]
 RC SEQUENCE FROM N.A.
 PC TISSUE=Colon carcinoma;
 RX MEDLINE=98326596; PubMed=9663463;
 Michel J.J., Xiong Y.;
 RA "Human CUL-1, but not other cullin family members, selectively
 RT interacts with SKP1 to form a complex with SKP2 and cyclin A.";
 RL Cell Growth Differ. 9:435-449(1998).
 RN [4]
 RC SEQUENCE OF 192-768 FROM N.A.
 RX MEDLINE=96279828; PubMed=8661378;
 Kipreos E.T., Lander L.E., Wing J.P., He W.W., Hedgecock E.M.;
 RT "cul-1 is required for cell cycle exit in C. elegans and identifies a

```
RT novel gene family." ;
RL Cell 85:829-839(1996).
RN [5]
RP SEQUENCE OF 426-768 FROM N.A.
RC TISSUE=Brain;
RA Yu W., Sarginson J., Gibbs R.A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CULLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF064087; AAC36304.1; -
DR EMBL; AB014517; BAA31592.1; -
DR EMBL; AF062537; AAC36682.1; -
DR EMBL; U58089; AAC50546.1; -
DR EMBL; AF052147; AAC28621.1; -
DR MIM; 603136; -
DR InterPro; IPR001373; Cullin.
DR Pfam; PF00888; Cullin; 1.
DR SMART; SM00182; CULLIN; 1.
DR PROSITE; PS01256; CULLIN_1; 1.
DR PROSITE; PS50069; CULLIN_2; 1.
FT CONFLICT 13 13 D -> G (IN REF. 3).
FT CONFLICT 159 179 DHLRQTLDDMIARERKGEVVD -> GSSTANSIGYDCKRAE
FT CONFLICT 426 451 RRSRS (IN REF. 3).
FT DVFERYKQHLARLLTNKSVSDDE -> MYLNVINNTW
FT QGDFSQIKVFLMTLK (IN REF. 5).
SQ SEQUENCE 768 AA; 89930 MW; ALA02022480BF099 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 768;
Best Local Similarity 66.7%; Pred. No. 47;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRIMHC 7
Db 246 ERVMHC 251
```

Search completed: November 1, 2002, 14:47:46
Job time : 2.36013 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 14:43:58 ; Search time 0.742765 Seconds
(without alignments)
905.569 Million cell updates/sec

Title: US-09-606-129a-18
Perfect score: 41
Sequence: 1 KKRIMHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	295	2 A42268	billiverdin reducta
2	41	100.0	633	2 T04537	hypothetical prote
3	38	92.7	296	2 G02086	billiverdin reducta
4	38	92.7	296	2 S62624	billiverdin reducta
5	36	87.8	512	2 H86206	hypothetical prote
6	34	82.9	401	2 F85015	probable phosphati
7	34	82.9	431	2 T01723	l-phosphatidylinos
8	33	80.5	189	2 A38117	hypothetical prote
9	33	80.5	189	2 AD2414	transposase all1486
10	33	80.5	189	2 AE2064	transposase all1206
11	33	80.5	189	2 AF2052	transposase all197
12	33	80.5	189	2 AF2152	transposase alr277
13	33	80.5	189	2 AG1875	transposase alr055
14	33	80.5	189	2 AG2037	transposase alr185
15	33	80.5	508	2 T02486	hypothetical prote
16	33	80.5	925	2 S50490	hypothetical prote
17	32	78.0	107	2 B64845	hypothetical prote
18	32	78.0	107	2 A99791	hypothetical prote
19	32	78.0	107	2 D85651	hypothetical prote
20	32	78.0	140	2 I49688	binding protein -
21	32	78.0	142	2 JC1365	FK506/rapamycin-bi
22	32	78.0	328	2 B65086	hydrogenase (EC 1.
23	32	78.0	328	2 A91114	hydrogenase-2 smal
24	32	78.0	328	2 A85959	hydrogenase-2 smal
25	32	78.0	328	2 AE0885	hydrogenase-2 smal
26	32	78.0	377	2 AB2213	methylocitrate synt
27	32	78.0	497	2 C86463	hypothetical prote
28	32	78.0	516	2 T23827	hypothetical prote
29	32	78.0	521	2 C88827	protein M7.2 [mpo

ALIGNMENTS

RESULT 1

A42268
billiverdin reductase (EC 1.3.1.24) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A42268
R:Fakhrai, H.; Maines, M.D.
J. Biol. Chem. 267, 4023-4029, 1992
A:Title: Expression and characterization of a cDNA for rat kidney billiverdin reductas
A:Reference number: A42268; MUID:92156147
A:Accession: A42268
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid; protein
A:Residues: 1-295 <FAK>
A:Cross-references: GB:M81681; NID:g203177; PIDN:AAA40830.1; PID:g203178
A:Experimental source: kidney
A:Note: sequence extracted from NCBI backbone (NCBIP:82800)
C:Keywords: liver; oxidoreductase

Query Match 100.0%; Score 41; DB 2; Length 295;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
Db 274 KKRIMHC 280

RESULT 2

T04537
hypothetical protein F28J12.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C:Accession: T04537
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, February 1998
A:Reference number: X15377
A:Accession: T04537
A:Molecule type: DNA
A:Residues: 1-633 <BEV>
A:Cross-references: EMBL:AL021710
A:Experimental source: cultivar Columbia; BAC clone F28J12
C:Genetics:
A:Map position: 4
A:Introns: 281/3; 303/3; 442/1; 614/3
A:Note: F28J12.70

Query Match 100.0%; Score 41; DB 2; Length 633;
Best Local Similarity 100.0%; Pred. No. 0.95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7

leucyl aminopeptid
leucyl aminopeptid
leucyl aminopeptid
leucyl aminopeptid
leucyl aminopeptid
probable membrane
hypothetical prote
ribosomal protein
hypothetical prote
C6 protein - rabbi
hypothetical prote
hypothetical prote
cyclin delta-2 - A
hypothetical prote
hypothetical prote
interferon-induced

|||||||
Db 591 KKRIMHC 597

RESULT 3

G02066

billyverdin reductase (EC 1.3.1.24) - human

N:Alternate names: billyverdin IX-alpha reductase

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998

C:Accession: G02066

R:Komuro, A.; Tobe, T.; Nakano, Y.; Yamaguchi, T.; Tomita, M.

submitted to the EMBL Data Library, August 1995

A:Reference number: H00768

A:Accession: G02066

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-296 <KOM>

A:Cross-references: EMBL:U34877; NID:gl143231; PID:gl143232

C:Keywords: oxidoreductase

Query Match 92.7%; Score 38; DB 2; Length 296;

Best Local Similarity 85.7%; Pred. No. 2;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7

|||||||

Db 275 KKRILHC 281

RESULT 4

S62624

billyverdin reductase (EC 1.3.1.24) - human

N:Alternate names: billyverdin IX-alpha reductase

C:Species: Homo sapiens (man)

C:Date: 28-Oct-1996 #sequence_revision 09-May-1997 #text_change 21-Jul-2000

C:Accession: S62624; S62622; S29736

R:Maines, M.D.; Polevoda, B.V.; Huang, T.J.; McCoubrey Jr., W.K.

Eur. J. Biochem. 235, 372-381, 1996

A:Title: Human billyverdin IX-alpha reductase is a zinc-metalloprotein. Characterization

A:Reference number: S62622; MUID:96202961

A:Accession: S62624

A:Molecule type: mRNA

A:Residues: 1-296 <MAI>

A:Cross-references: EMBL:X93086; NID:gl246748; PIDN:CAA63635.1; PID:gl246749

A:Accession: S62622

A:Molecule type: protein

A:Residues: 3-24, 'X', 26-27, 'X', 29-36; 48-74; 228-234; 235-248 <MAF>

R:Maines, M.D.; Trakshel, G.M.

Arch. Biochem. Biophys. 300, 320-326, 1993

A:Title: Purification and characterization of human billyverdin reductase.

A:Reference number: S29736; MUID:93143333

A:Accession: S29736

A:Molecule type: protein

A:Residues: 3-24, 'X', 26-27, 'X', 29-36; 48-74; 228-234; 235-248 <MAW>

A:Note: the sequence of peptide 1 from page 323 seems not to belong to this protein

C:Genetics:

A:Gene: BVR

C:Keywords: oxidoreductase

F:3-296/Product: billyverdin reductase IX-alpha #status experimental <MAT>

Query Match 92.7%; Score 38; DB 2; Length 296;

Best Local Similarity 85.7%; Pred. No. 2;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7

|||||||

Db 275 KKRILHC 281

RESULT 5

H86206

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86206

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: H86206

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-512 <STO>

A:Cross-references: GB:AE005172; NID:g8954028; PIDN:AAF82202.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 87.8%; Score 36; DB 2; Length 512;

Best Local Similarity 71.4%; Pred. No. 8.7;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKRIMHC 7

|||||||

Db 236 KKRILHC 242

RESULT 6

F85015

probable phosphatidylinositol kinase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: F85015

R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: F85015

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-401 <STO>

A:Cross-references: GB:NC_001268; NID:g7267616; PIDN:CAB80928.1; GSPDB:GN00140

C:Genetics:

A:Gene: Atg01190

A:Map position: 4

Query Match 82.9%; Score 34; DB 2; Length 401;

Best Local Similarity 85.7%; Pred. No. 18;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIMHC 7

|||||||

Db 353 KKRILHC 359

RESULT 7

T01723

1-phosphatidylinositol-4-phosphate 5-kinase type II homolog - Arabidopsis thaliana

N:Alternate names: protein AIG002N01.9

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999

C:Accession: T01723

R:Scheet, P.; Maggi, L.

submitted to the EMBL Data Library, June 1997

A:Description: The sequence of A. thaliana IG002N01.

A:Reference number: Z14407

A:Accession: T01723

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB74472.1; PID:gl7131866; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2773

Query Match 80.5%; Score 33; DB 2; Length 189;
Best Local Similarity 71.4%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||: ||
Db 171 KKRLLHC 177

RESULT 13

AG1875
transposase alr0552 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AG1875
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1875
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB72510.1; PID:gl7129897; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0552

Query Match 80.5%; Score 33; DB 2; Length 189;
Best Local Similarity 71.4%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||: ||
Db 171 KKRLLHC 177

RESULT 14

AG2037
transposase alr1853 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AG2037
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB73552.1; PID:gl7130943; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1853

Query Match 80.5%; Score 33; DB 2; Length 189;
Best Local Similarity 71.4%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7
|||: ||
Db 171 KKRLLHC 177

RESULT 15

T02486
hypothetical protein At2g2990 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F23F1.9
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02486; B84703
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC F23F1 genomic sequence.
A:Reference number: 214675
A:Accession: T02486
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-508 <ROU>
A:Cross-references: EMBL:AC004680; NID:g3420043; PID:g3420052
A:Experimental source: Cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84703
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <STO>
A:Cross-references: GB:AE002093; NID:g3420052; PIDN:AAC31853.1; GSPDB:GN00139
C:Genetics:

A:Gene: At2g2990; F23F1.9

A:Map position: 2

A:Introns: 158/3; 230/1; 283/3; 305/3; 360/2; 398/3; 458/3

C:Superfamily: NADH dehydrogenase

Query Match 80.5%; Score 33; DB 2; Length 508;

Best Local Similarity 57.1%; Pred. No. 36;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKRIMHC 7

|||: ||

Db 236 KKRLLHC 242

Search completed: November 1, 2002, 14:50:57

Job time : 2.90943 secs